CORRECTED VERSION*

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12Q 1/68, C12P 19/34, C07H 21/04, C07K 14/435, G01N 33/53

A1

(11) International Publication Number: WO 98/24935

(43) International Publication Date: 11 June 1998 (11.06.98)

(21) International Application Number: PCT/US97/22105

(22) International Filing Date: 5 December 1997 (05.12.97)

(30) Priority Data:

60/032,619	6 December 1996 (06.12.96)	US
60/032,701	12 December 1996 (12.12.96)	US
60/041,576	24 March 1997 (24.03.97)	US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications

US	60/032,619 (CIP)
Filed on	6 December 1996 (06.12.96)
US	60/032,701 (CIP)
Filed on	12 December 1996 (12.12.96)
US	60/041,576 (CIP)
Filed on	24 March 1997 (24.03.97)

(71) Applicant *(for all designated States except US)*: UROCOR, INC. [US/US]; 800 Research Parkway, Oklahoma City, OK 73104 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: RALPH, David [US/US]; 2504 Stonehenge Drive, Edmond, OK 73034 (US). AN, Gang [CN/US]; 2609 N.W. 160th Street, Edmond, OK 73013 (US). O'HARA, Mark [US/US]; 130 N.W. 17th Street, Oklahoma City, OK 73101 (US). VELTRI, Robert [US/US]; 7301 N.E. 106th Street, Oklahoma City, OK 73151 (US).

(74) Agent: NAKASHIMA, Richard, A.; Arnold, White & Durkee, P.O. Box 4433, Houston, TX 77210 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: DIAGNOSIS OF DISEASE STATE USING MRNA PROFILES

(57) Abstract

Disclosed are diagnostic techniques for the detection of a human diseased state. Genetic probes and methods useful in monitoring the progression and diagnosis of the disease state are described. The invention relates particularly to probes and methods for evaluating the presence of RNA species that are differentially expressed in the peripheral blood of individuals with the disease state compared to normal healthy individuals. Further disclosed is a multivariate diagnostic model for prostate cancer in a population of men with moderately elevated total serum PSA (≥ 2.0 ng/ml). Results of quantitative serum assays for the UC325 gene product [Interleukin-8 (IL-8)], total prostate specific antigen (t-PSA), as well as Free/Total (f/t PSA) ratios were combined to enhance the sensitivity of prostate cancer diagnosis in a defined urologic population diagnosed either organ-confined prostate cancer (clinical stage A & B), non-organ-confined prostate cancer (clinical stage C or D) or benign prostatic hyperplasia (BPH). The additional ability of UC325 gene product serum levels to accurately stage prostate cancer independently of t-PSA of f/t PSA is disclosed.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12Q 1/68, C12P 19/34, C07H 21/04, C07K 14/435, G01N 33/53

A1

(11) International Publication Number: WO 98/24935

(43) International Publication Date: 11 June 1998 (11.06.98)

(21) International Application Number: PCT/US97/22105

(22) International Filing Date: 5 December 1997 (05.12.97)

(30) Priority Data:

60/032,619	6 December 1996 (06.12.96)	US
60/032,701	12 December 1996 (12.12.96)	US
60/041,576	24 March 1997 (24.03.97)	US

(71) Applicant *(for all designated States except US)*: UROCOR, INC. [US/US]; 800 Research Parkway, Oklahoma City, OK 73104 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: RALPH, David [US/US]; 2504 Stonehenge Drive, Edmond, OK 73034 (US). AN, Gang [CN/US]; 2609 N.W. 160th Street, Edmond, OK 73013 (US). O'HARA, Mark [US/US]; 130 N.W. 17th Street, Oklahoma City, OK 73101 (US). VELTRI, Robert [US/US]; 7301 N.E. 106th Street, Oklahoma City, OK 73151 (US).

(74) Agent: CORDER, Timothy, S.; Arnold, White & Durkee, P.O. Box 4433, Houston, TX 77210 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: DIAGNOSIS OF DISEASE STATE USING MRNA PROFILES

(57) Abstract

Disclosed are diagnostic techniques for the detection of a human diseased state. Genetic probes and methods useful in monitoring the progression and diagnosis of the disease state are described. The invention relates particularly to probes and methods for evaluating the presence of RNA species that are differentially expressed in the peripheral blood of individuals with the disease state compared to normal healthy individuals. Further disclosed is a multivariate diagnostic model for prostate cancer in a population of men with moderately elevated total serum PSA (≥ 2.0 ng/ml). Results of quantitative serum assays for the UC325 gene product [Interleukin-8 (IL-8)], total prostate specific antigen (t-PSA), as well as Free/Total (f/t PSA) ratios were combined to enhance the sensitivity of prostate cancer diagnosis in a defined urologic population diagnosed either organ-confined prostate cancer (clinical stage A & B), non-organ-confined prostate cancer (clinical stage C or D) or benign prostatic hyperplasia (BPH). The additional ability of UC325 gene product serum levels to accurately stage prostate cancer independently of t-PSA of f/t PSA is disclosed.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

DESCRIPTION

DIAGNOSIS OF DISEASE STATE USING MRNA PROFILES

BACKGROUND OF THE INVENTION

A. Field of the Invention

The present invention relates generally to the detection and diagnosis of human disease states and methods relating thereto. More particularly, the present invention concerns probes and methods useful in diagnosing, identifying and monitoring the progression of disease states through measurements of gene products.

B. Description of the Related Art

Genetic detection of human disease states is a rapidly developing field (Taparowsky *et al.*, 1982; Slamon *et al.*, 1989; Sidransky *et al.*, 1992; Miki *et al.*, 1994; Dong *et al.*, 1995; Morahan *et al.*, 1996; Lifton, 1996; Barinaga, 1996). One advantage presented by this field is that certain disease states may be detected by non-invasive means, *e.g.* sampling peripheral blood or amniotic fluid. Affected individuals may be diagnosed early in disease progression, allowing more effective patient management with better clinical outcomes.

Some problems exist with this approach. A number of known genetic lesions merely predispose to development of specific disease states. Individuals carrying the genetic lesion may not develop the disease state, while other individuals may develop the disease state without possessing a particular genetic lesion. In human cancers, genetic defects may potentially occur in a large number of known tumor suppresser genes and proto-oncogenes.

The genetic detection of cancer has a long history. One of the earliest genetic lesions shown to predispose to cancer was transforming point mutations in the *ras* oncogenes (Taparowsky *et al.*, 1982). Transforming *ras* point mutations may be detected in the stool of individuals with benign and malignant colorectal tumors (Sidransky *et al.*, 1992). However, only 50% of such tumors contained a *ras* mutation (Sidransky *et al.*, 1992). Similar results have been obtained with amplification of HER-2/*neu* in breast and ovarian cancer (Slamon *et al.*, 1989), deletion and

mutation of p53 in bladder cancer (Sidransky *et al.*, 1991), deletion of DCC in colorectal cancer (Fearon *et al.*, 1990) and mutation of *BRCA1* in breast and ovarian cancer (Miki *et al.*, 1994).

None of these genetic lesions are capable of predicting a majority of individuals with cancer and most require direct sampling of a suspected tumor, making screening difficult. Further, none of the markers described above are capable of distinguishing between metastatic and non-metastatic forms of cancer. In effective management of cancer patients, identification of those individuals whose tumors have already metastasized or are likely to metastasize is critical. Because metastatic cancer kills 560,000 people in the US each year (ACS home page), identification of markers for metastatic cancer, such as metastatic prostate and breast cancer, would be an important advance.

A particular problem in cancer detection and diagnosis occurs with prostate cancer. Carcinoma of the prostate (PCA) is the second-most frequent cause of male cancer-related death in the United States (Boring, 1993). The incidence of prostate cancer increased by 50% between 1980 and 1990 (Stone *et al.*, 1994). Although relatively few prostate tumors progress to clinical significance during the lifetime of the patient, those which are progressive in nature are likely to have metastasized by the time of detection. Survival rates for individuals with metastatic prostate cancer are quite low. Between these extremes are patients with prostate tumors that will metastasize but have not yet done so, for whom surgical prostate removal is curative. Determination of which group a patient falls within is critical in determining optimal treatment and patient survival.

Genetic changes reported to be associated with prostate cancer include: allelic loss (Bova, *et al.*, 1993; Macoska *et al.*, 1994; Carter *et al.*, 1990); DNA hypermethylation (Isaacs *et al.*, 1994); point mutations or deletions of the retinoblastoma (Rb) and p53 genes (Bookstein *et al.*, 1990a; Bookstein *et al.*, 1990b; Isaacs *et al.*, 1991); and aneuploidy and aneusomy of chromosomes detected by fluorescence in situ hybridization (FISH) (Macoska *et al.*, 1994; Visakorpi *et al.*, 1994; Takahashi *et al.*, 1994; Alcaraz *et al.*, 1994).

A recent development in this field was the identification of a prostate metastasis suppresser gene, *KAI1* (Dong *et al.*, 1995). Insertion of wild-type *KAI1* gene into a rat prostate cancer line caused a significant decrease in metastatic tumor formation (Dong *et al.*, 1995). However, detection of *KAI1* mutations is dependent upon direct sampling of mutant prostate cells. Thus, either a primary prostate tumor must be sampled or else sufficient transformed cells must be present

in blood, lymph nodes or other tissues to detect the missing or abnormal gene. Further, the presence of a deleted gene may frequently be masked by large numbers of untransformed cells that may be present in a given tissue sample.

The most commonly utilized current tests for prostate cancer are digital rectal examination (DRE) and analysis of serum prostate specific antigen (PSA). Although PSA has been widely used as a clinical marker of prostate cancer since 1988 (Partin & Oesterling, 1994), screening programs utilizing PSA alone or in combination with digital rectal examination have not been successful in improving the survival rate for men with prostate cancer (Partin & Oesterling, 1994). While PSA is specific to prostate tissue, it is produced by normal and benign as well as malignant prostatic epithelium, resulting in a high false-positive rate for prostate cancer detection (Partin & Oesterling, 1994).

Other markers that have been used for prostate cancer detection include prostatic acid phosphatase (PAP) and prostate secreted protein (PSP). PAP is secreted by prostate cells under hormonal control (Brawn *et al.*, 1996). It has less specificity and sensitivity than does PSA. As a result, it is used much less now, although PAP may still have some applications for monitoring metastatic patients that have failed primary treatments. In general, PSP is a more sensitive biomarker than PAP, but is not as sensitive as PSA (Huang *et al.*, 1993). Like PSA, PSP levels are frequently elevated in patients with BPH as well as those with prostate cancer.

Another serum marker associated with prostate disease is prostate specific membrane antigen (PSMA) (Horoszewicz *et al.*, 1987; Carter *et al.*, 1996; Murphy *et al.*, 1996). PSMA is a Type II cell membrane protein and has been identified as Folic Acid Hydrolase (FAH) (Heston, 1996; Carter *et al.*, 1996). Antibodies against PSMA react with both normal prostate tissue and prostate cancer tissue (Horoszewicz *et al.*, 1987). Murphy *et al.* (1995) used ELISA to detect serum PSMA in advanced prostate cancer. As a serum test, PSMA levels are a relatively poor indicator of prostate cancer. However, PSMA may have utility in certain circumstances. PSMA is expressed in metastatic prostate tumor capillary beds (Silver *et al.*, 1997) and is reported to be more abundant in the blood of metastatic cancer patients (Murphy *et al.*, 1996). PSMA messenger RNA (mRNA) is down-regulated 8-10 fold in the LNCaP prostate cancer cell line after exposure to 5-α-dihydroxytestosterone (DHT) (Israeli *et al.*, 1994).

A relatively new potential biomarker for prostate cancer is human kallekrein 2 (HK2) (Piironen *et al.*, 1996). HK2 is a member of the kallekrein family that is secreted by the prostate gland. In theory, serum concentrations of HK2 may be of utility in prostate cancer detection or diagnosis, but the usefulness of this marker is still being evaluated.

Interleukin 8 (IL-8) is a potent serum cytokine that is synthesized and secreted by a large variety of cell types, including neutrophils, endothelial cells, T-cells, macrophages, monocytes, and fibroblasts (Saito *et al.*, 1994). Previous reports have found overexpression of IL-8 in some forms of cancer. (di Celle *et al.*, 1994; Ikei *et al.*, 1992; Scheibenbogen *et al.*, 1995; Vinante *et al.*, 1993). RT-PCR analysis was used by di Celle *et al.* (1994) to demonstrate IL-8 production in B-cell chronic lymphocytic leukemia. Vinante *et al.* (1993) used Northern blot analysis to show upregulation of IL-8 expression in acute myelogenous leukemia. Ikei *et al.* (1992) found an increase in serum levels of IL-8 in hepatic cancer patients following therapeutic treatment. Scheibenbogen *et al.* (1995) observed a correlation between IL-8 levels and tumor loads in patients with metastatic melanoma, while reporting that serum IL-8 was undetectable in healthy individuals or in patients with metastatic renal cell carcinoma. These authors suggested that the IL-8 was produced by the melanoma cells themselves, rather than by circulating lymphocytes. Andrawis *et al.* (1996) reported that while IL-8 was expressed in prostate and bladder cancer, it was also abundantly expressed in normal bladder epithelium and in some basal cells in BPH.

The instant disclosure is the first to combine measurement of IL-8 gene products with serum markers of prostate disease, such as PSA, PAP, HK2 or PSMA. The surprising result of this multivariate detection is a dramatic increase in sensitivity and specificity of prostate cancer detection, while simultaneously allowing the differentiation of advanced from localized forms of prostate tumor.

SUMMARY OF THE INVENTION

The present invention addresses deficiencies in the prior art by providing methods for identifying specific disease state markers that are expressed in peripheral lymphocytes of patients in response to a disease state, at a different level than such markers are expressed in the peripheral blood of a normal subject (a healthy individual). An important advantage provided by the present

invention is that a disease state may be detected, diagnosed, or a prognosis may be derived by examining a blood sample rather than relying on a more invasive, or less sensitive test. In addition, a subject may be monitored for disease progression, status and response to therapies through monitoring of differentially expressed disease markers. In certain embodiments of the present invention a "patient" "individual" or "subject" may be an animal, including a laboratory animal or other animal species, or in certain embodiments a human subject.

In certain embodiments of the invention the terms "expression", "gene expression" and "expression products" may refer to either production of a marker gene RNA message or the RNA message produced or both. In certain other embodiments of the invention the terms "expression", "gene expression" and "expression products" may refer to either translation of a marker RNA message into proteins, polypeptides and/or peptides, and/or to the produced proteins, polypeptides, and/or peptides. In certain aspects of the invention a marker may be a gene whose expression is activated to a higher level in a patient suffering a disease state, relative to its expression in a healthy subject. It is also understood that a differentially expressed marker may be either activated or inhibited at the nucleic acid level or protein level, or may it may subject to alternative splicing to result in a different polypeptide product. Such differences may be evidenced by a change in mRNA levels, surface expression, secretion or other partitioning of a polypeptide, for example. In certain aspects of the invention, a marker may be a comparison of expression between two or more marker genes, and/or a comparison of the ratios of the expression between two or more marker genes, or even a comparison of two differently processed products of the same gene, which differ between healthy subjects and subjects suffering a diseased state.

As demonstrated in the examples included herein, the present inventors have identified certain markers and methods of identifying markers that have been applied for the detection of metastatic prostate and metastatic breast cancer. These examples have demonstrated that disease states may be detected and monitored by surveying the response of healthy immune cells to the disease condition. As such, this novel method is contemplated to be suitable for detection of markers that are differentially expressed in response to other forms of cancer as well as other diseases such as asthma, lupus erythematosis, rheumatoid arthritis, multiple sclerosis, myasthenia

gravis, autoimmune thyroiditis, amyloid lateral sclerosis, interstitial cystitis, prostatitis or other systemic or chronic conditions.

In a certain embodiment of the present invention, the inventors have demonstrated the ability to detect and discriminate between benign prostatic hyperplasia (BPH) and prostate cancer, using multivariate analysis with several different prostate disease markers. By combining test results for serum prostate specific antigen (PSA) and IL-8 gene products, it is possible to identify a significant proportion of individuals with prostate cancer, while achieving close to one hundred percent accuracy in differentiating between individuals presenting with prostate cancer versus BPH. These levels of sensitivity and specificity represent significant advances over the prior art in prostate cancer detection and differentiation, which traditionally have been performed with univariate analysis with PSA, digital rectal examination and other techniques. It is further disclosed that levels of IL-8 gene product in the peripheral circulation may be used to discriminate advanced from localized stages of prostate cancer.

It is an important aspect of the present invention that it is the response of the normal blood lymphocytes that is being examined, rather than the prostate, breast or other disease cells themselves as in previous methods. As an aspect of the invention, certain mRNAs are identified that are differentially expressed in normal cells, as a reaction to a disease state, relative to their expression in healthy subjects. Two of the metastatic cancer-markers disclosed herein represent previously unreported genes, with one of the two matching a small expressed sequence tag (EST) described in Genebank Accession # T03013 and SEQ ID NO:1, and another matching the sequence disclosed in SEQ ID NO:2. Another marker corresponds to the sequence of elongation factor 1-alpha (Genebank Accession # X03558 and SEQ ID NO:3). Two other markers represent alternatively spliced forms (Genebank Accession # M28130 and SEQ ID NO:5; Genebank Accession # Y00787 and SEQ ID NO:4) of mRNA from the IL-8 (interleukin 8) gene. One metastatic cancer marker is a previously uncharacterized gene (SEQ ID NO:29) that has homology to a number of previously identified EST sequences, while another marker is a previously identified gene sequence (KA000262, Genebank Accession # D87451).

The markers and marker genes comprising the group of total prostate specific antigen (PSA); prostate specific membrane antigen (PSMA=Folic Acid Hydrolase); prostate acid phosphatase (PAP); prostatic secretory proteins (PSP_{94}); human kallekrein 2 (HK2); and the ratio

of the concentrations of free and bound forms of PSA (f/t PSA), in combination with any of the markers identified herein as SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:4, SEQ ID NO:2 or SEQ ID NO:29, and the sequences identified in Genebank Accession #s D87451, T03013, X03558, M28130, and Y00787, their complementary nucleic acid sequences, or their expression products may be used in all embodiments using or detecting the markers in any of the methods disclosed herein or known in the art. In the examples disclosed herein, the differential expression of marker genes is detected by RNA fingerprinting methods, however, differential expression detected by any other means, including but not limited to other RNA fingerprinting methods, Northern blotting, immunodetection, protein-protein interactions, biological activity or other methods known in the art would fall within the scope of the present invention.

The present disclosure is the first report of an alternatively spliced form of IL-8 mRNA that includes intron 3. In the peripheral blood of normal individuals the mRNA transcript containing intron 3 (Genebank Accession # M28130) is more abundant than the previously reported spliced form from which intron 3 is missing (Genebank Accession # Y00787). Surprisingly, in patients with metastatic prostate cancer the previously reported spliced form is much more abundant, with a seven-fold increase compared to normal individuals. In contrast, the transcript containing intron 3 is approximately seven-fold less abundant in patients with metastatic prostate cancer than in normal individuals.

The substantial change in levels of alternatively spliced mRNA species in the peripheral blood of individuals with metastatic cancer provides a simple and effective diagnostic test for the presence of cancer metastases, that is unaffected by problems in sampling primary tumors or the masking influence of normal cells in a tissue sample. It therefore represents a significant advance over previous methods for detecting and diagnosing metastatic cancer in humans. The skilled practitioner will realize that metastatic cancer detection and diagnosis may be performed by quantitative analysis of either the IL-8 mRNA transcripts themselves or their protein products.

The present disclosure represents a substantial and unexpected advance over previous knowledge in this field. It reports a novel spliced form of IL-8 mRNA that is repressed in metastatic prostate cancer. It provides a sensitive means for detecting metastatic cancer by measuring the levels of the two alternatively spliced IL-8 mRNA forms. It provides a highly sensitive and specific method for detecting and differentiating between BPH, localized and

advanced forms of prostate cancer by combining detection of IL-8 gene product with other markers of prostate disease.

The present disclosure further demonstrates the feasibility of detecting and diagnosing human disease states in general by monitoring changes in the expression of specific genes in peripheral lymphocytes. The skilled practitioner of the art will realize that such a technique has widespread applicability for screening of asymptomatic individuals for disease state markers.

The identified disease state markers may in turn be used to design specific oligonucleotide probes and primers. In certain preferred embodiments the term "primer" as used here includes any nucleic acid capable of priming template-dependent synthesis of a nascent nucleic acid. In certain other embodiments the nucleic acid may be able to hybridize a template, but not be extended for synthesis of nascent nucleic acid that is complementary to the template. As used herein a "primer" may be at least about 5, about 6, about 7, about 8, about 9, about 10, about 11, about 12, about 13, about 14, about 15, about 16, about 17, about 18, about 19, about 20, about 21, about 22, about 23, about 24, about 25, about 26, about 27, about 28, about 29, about 30, about 35, about 40, about 50, about, 75, about 100, about 150, about 200, about 300, about 400, about 500, to one base shorter in length than the template sequence at the 3' end of the primer to allow extension of a nascent nucleic acid chain, though the 5' end of the primer may extend in length beyond the 3' end of the template sequence. In certain embodiments of the present invention the term "template" may refer to a nucleic acid that is used in the creation of a complementary nucleic acid strand to the "template" strand. The template may be either RNA and/or DNA, and the complementary strand may also be RNA and/or DNA. In certain embodiments the complementary strand may comprise all or part of the complementary sequence to the template, and/or may include mutations so that it is not an exact, complementary strand to the template. Strands that are not exactly complementary to the template strand may hybridize specifically to the template strand in detection assays described here, as well as other assays known in the art, and such complementary strands that can be used in detection assays are part of the invention.

When used in combination with nucleic acid amplification procedures, these probes and primers enable the rapid analysis of peripheral blood samples. In certain aspects of the invention, the term "amplification" may refer to any method or technique known in the art or described herein for duplicating or increasing the number of copies or amount of a target nucleic acid or its

complement. In certain aspects of the invention, the term "amplicon" refers to the target sequence for amplification, or that part of a target sequence that is amplified, and/or the amplification products of the target sequence being amplified. In certain other embodiments an "amplicon" may include the sequence of probes or primers used in amplification. This analysis assists physicians in detecting and diagnosing the disease state and in determining optimal treatment courses for individuals at varying stages of disease state progression.

In light of the present disclosure, one of ordinary skill in the art will select segments from the identified marker genes for use in the different detection, diagnostic, or prognostic methods, vector constructs, antibody production, kit, and/or any of the embodiments described herein as part of the present invention. Marker gene sequences include those published in the Genebank database that match the identified marker genes: Genebank Accession numbers D87451, T03013, X03558, M28130, and Y00787, as well as the sequences disclosed herein as SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, and SEQ ID NO:29, which also include sequences for previously uncharacterized marker genes (UCPB 35, SEQ ID NO:1; UC 302, SEQ ID NO:2; UC 331, SEQ ID NO:29) identified herein. For example, in certain embodiments in which one may be practicing the present invention for the identification of a disease marker, for example, the sequences selected to design probes and primers may include repetitive stretches of adenine nucleotides (poly-A tails) normally attached at the ends of the RNA for the identified marker genes. In certain other embodiments, probes and primers may be specifically designed to not include these or other segments from the identified marker genes, as one of ordinary skilled in the art may deem certain segments more suitable for use in the detection methods disclosed.

For example, where a genomic sequence is disclosed, one would use sequences that correspond to exon regions of the gene in most cases. However, as described herein, at least one metastatic cancer marker includes alternately spliced transcripts so that intronic sequences may be used for diagnostic or prognostic purposes (Genebank Accession # M28130). Exon sequences in the gene structure, as described in the Genebank listing for Accession # M28130, include bases 1482 to 1647, 2464 to 2599, 2871 to 2954, and 3370 to 4236. Intron 3 includes bases 2954 to 3370. One of ordinary skill in the art may select segments from the published exon sequences, or may assemble them into a reconstructed mRNA sequence that does not contain intronic sequences, such as intron 3. Alternatively, the published sequence for IL-8 that reports a spliced form from

which intron 3 is missing (Genebank Accession # Y00787) may be used. By choosing or selecting the sequences to include or exclude intron 3, one could preferentially detect expression of one alternatively spliced form of IL-8, or even the ratio of the two forms using the methods disclosed herein. One of ordinary skill in the art may select and/or assemble segments from any of the identified marker gene sequences into other useful forms, such as coding segment reconstruction's of mRNA sequences from published genomic sequences of the identified marker genes, as part of the present invention. Such assembled sequences would be useful in designing probes and primers, as well as providing coding segments for protein translation, for detection, diagnosis, and prognosis embodiments of the invention described herein.

For example, primers to detect the message of IL 8 using the transcribed portions of the marker sequence as set forth in the listing in Genebank Accession # M28130 may hybridize to nucleotides 1482 to 1503 and the complement of nucleotides 1626-1647. These particular primers would amplify a segment of message of the marker gene 166 base pairs in length. Primers designed to nucleotides 1482 to 1503 and the complement of nucleotides 2464 to 2483 would amplify a segment of message of the marker gene 186 base pairs long in messages that have the intervening intron between nucleotides 1648 to 2463 removed. Thus, one skilled in the art would be able to calculate the expected size of transcribed sequences from marker genes identified herein whose sequences are published either as genomic sequence, mRNA, or cDNA, as well as the sequences disclosed herein, taking into account the differences in size of the products produced depending on the presence or absence of intronic sequences. In preferred embodiments, the differences in size of amplification products using primers designed to regions flanking both sides of intron 3 in the IL-8 marker gene sequences identified (Genebank Accession # Y00787 and # M28130) can be used in detection, diagnosis, and/or prognosis of metastatic cancer. However, primers designed to regions of IL-8 sequences that do not flank intron 3, or the other marker genes that do not have differences in intron splicing, or that prime mRNA or cDNA template sequences, would not be expected to produce amplification products that include intronic segments.

For example, primers designed to nucleotides 1 to 20 and the complement of nucleotides 200 to 220 of SEQ ID NO:1 would amplify a metastatic marker gene segment 220 base pairs long. Primers designed to nucleotides 115 to 138 and the complement of nucleotides 730 to 744 of SEQ ID NO:29 would amplify a metastatic marker gene segment 630 base pairs long. Primers designed

to nucleotides 102 to 120 and the complement of nucleotides 381 to 401 of the IL-8 marker gene sequence identified in Genebank Accession # Y00787 would amplify a metastatic marker gene segment 302 base pairs long that would be approximately sevenfold less abundant in normal patients when compared to patients with metastatic prostate cancer. Primers can be designed to amplify the transcribed portions of the metastatic cancer markers that would include any length of nucleotide segment of the transcribed sequences, up to and including the full length of each marker gene message. It is preferred that the amplified segments of identified marker genes be an amplicon of at least about 50 to about 500 base pairs in length. It is particularly preferred that the amplified segments of identified marker genes be an amplicon of at least about 100 to about 415 base pairs in length, and/or no longer in length than the amplified segment used to normalize the quantity of message being amplified in the detection assays described herein. Such assays include RNA fingerprinting methods, however, differential expression may be detected by other means, and all such methods would fall within the scope of the present invention. The predicted size of the amplified metastatic cancer marker gene segment, calculated by the location of the primers relative to the transcribed sequence, would be used to determine if the detected amplification product is indeed the marker gene being amplified. Sequencing the amplified or detected band that matches the expected size of the amplification product and comparison of the band's sequence to the known or disclosed sequence of the marker gene would confirm that the correct marker gene is being amplified and detected.

The identified markers may also be used to identify and isolate full length gene sequences, including regulatory elements for gene expression, from genomic human DNA libraries. The cDNA sequences identified in the present disclosure may be used as hybridization probes to screen genomic human DNA libraries by conventional techniques. Once partial genomic clones have been identified, full-length genes may be isolated by "chromosomal walking" (also called "overlap hybridization"). *See,* Chinault & Carbon "Overlap Hybridization Screening: Isolation and Characterization of Overlapping DNA Fragments Surrounding the *LEU2* Gene on Yeast Chromosome III." *Gene* 5: 111-126, 1979. Once a partial genomic clone has been isolated using a cDNA hybridization probe, nonrepetitive segments at or near the ends of the partial genomic clone may be used as hybridization probes in further genomic library screening, ultimately allowing isolation of entire gene sequences for the disease state markers of interest. It will be recognized that

full length genes may be obtained using the small expressed sequence tags (ESTs) described in this disclosure using technology currently available and described in this disclosure (Sambrook *et al.*, 1989; Chinault & Carbon, 1979). Sequences identified and isolated by such means may be useful in the detection of the prostate marker genes using the detection methods described herein, and are part of the invention.

The identified markers may be used to identify and isolate cDNA sequences. The EST sequences identified in the present disclosure may be used as hybridization probes to screen human cDNA libraries by conventional techniques. It will be recognized that these techniques would start by obtaining a high quality human cDNA library, many of which are readily available from commercial or other sources. The library may be plated on, for example, agarose plates containing nutrients, antibiotics and other conventional ingredients. Individual colonies may then be transferred to nylon or nitrocellulose membranes and the EST probes hybridized to complementary sequences on the membranes. Hybridization may be detected by radioactive or enzyme-linked tags associated with the hybridized probes. Positive colonies may be grown up and sequenced by, for example, Sanger dideoxynucleotide sequencing or similar methods well known in the art. Comparison of cloned cDNA sequences with known human or animal cDNA or genomic sequences may be performed using computer programs and databases well known in the art. Sequences identified and isolated by such means may be useful in the detection of the prostate disease, or other disease marker genes using the detection methods described herein, and are part of the invention.

In one embodiment of the present invention, the isolated nucleic acids of the identified marker genes are incorporated into expression vectors and expressed as the encoded proteins or peptides. Isolated nucleic acid segments may be from published sequences identified, or the sequences disclosed herein, as marker genes. Coding sequences may be assembled from amino acid encoding segments of marker genes to remove noncoding segments, or to truncate coding sequence, or to use the coding sequences or segments thereof in expression vectors as is known in the art. In certain embodiments, genomic sequences may be used to express peptides or proteins of the metastatic cancer maker genes identified herein.

Such proteins or peptides are in turn used as antigens for induction of monoclonal or polyclonal antibody production. Such antibodies may in turn be used to detect expressed proteins

as additional markers for human disease states. Antibody-protein binding may be detected and quantitated by a variety of means known in the art, such as labeling with fluorescent or radioactive ligands.

Certain metastatic marker genes disclosed herein (SEQ ID NO:1 and Genebank accession # T03013; and SEQ ID NO:2) do not have reading frames for translation disclosed. However, one of ordinary skill in the art may translate the identified sequences or segments thereof in the three potential reading frames to obtain peptides or proteins for use in generating antibodies to these marker genes. Such antibodies may be used to purify the proteins of the marker genes, and the identity of protein being detected is confirmed by peptide sequencing of the protein. Once confirmed as binding the translation products of the marker genes corresponding to SEQ ID NO:1 and Genebank accession # T03013, and/or SEQ ID NO:2, the antibodies that bind the marker gene protein would be useful in detecting, diagnosis, or prognosis of metastatic cancer.

An example of an marker gene sequence that would be preferred for translation would be intron 3 of IL-8 (Genebank Accession # M28130). Peptides or polypeptides that contain amino acid sequences from this intron would be preferred in the creation of polyclonal or monoclonal antibodies that preferentially detect forms of IL-8 which include intron 3.

In certain aspects of the present invention the terms "immunodetection", "immunobinding", "immunoreaction", "immunohistochemical", "immunosorbent", and "radioimmunoassays" refers to methods that concern binding, purifying, removing, quantifying or otherwise generally detecting biological components by obtaining a sample suspected of containing a protein, peptide or antibody, and contacting the sample with an antibody or protein or peptide in accordance with the present invention, as the case may be, under conditions effective to allow the formation of immunocomplexes. In certain preferred aspects of the present invention, one obtains a sample suspected of containing a disease state-marker encoded protein, peptide or a corresponding antibody, and contacts the sample with an antibody or encoded protein or peptide, as the case may be, and then detects or quantifies the amount of immune complex formed under the specific conditions. The steps of various useful immunodetection methods have been described in the scientific literature, such as, e.g., Nakamura *et al.* (1987).

In another embodiment of the present invention, the aforementioned oligonucleotide hybridization probes and primers are specific for disease state markers comprising isolated nucleic

acids of a sequence comprising the sequences published in Genebank Accession numbers D87451, T03013, X03558, M28130, and Y00787, as well as the sequences disclosed herein as SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, and SEQ ID NO:29. Such probes and primers may be of any length that would specifically hybridize to the identified marker gene sequences and may be at least about 14, about 15, about 16, about 17, about 18, about 19, about 20, about 21, about 22, about 23, about 24, about 25, about 26, about 27, about 28, about 29, about 30, about 35, about 40, about 50, about, 75, about 100, about 150, about 200, about 300, about 400, about 500, and in the case of probes, up to the full length of the sequences of the marker genes identified herein. Probes may also include additional sequence at their 5' and/or 3' ends so that they extend beyond the target sequence with which they hybridize. Such primers may be used to amplify disease state markers present in a biological sample, such as peripheral human blood. Amplification increases the sensitivity of various known techniques for detecting the presence of nucleic acid markers for human disease. Probes that hybridize with nucleic acid markers for human disease may be detected by conventional labeling methods, such as binding of fluorescent or radioactive ligands. The availability of probes and primers specific for such unique markers provides the basis for diagnostic kits identifying disease state progression.

An embodiment of the present invention encompasses a kit for detecting a disease state in a biological sample, comprising pairs of primers for amplifying nucleic acids corresponding to the marker genes and containers for each of these primers. In another embodiment, the invention encompasses a kit for detecting a disease state in a biological sample, comprising oligonucleotide probes that bind with high affinity to markers of the disease state and containers for each of these probes. In a further embodiment, the invention encompasses a kit for detecting a disease state in a biological sample, comprising antibodies specific for proteins encoded by the nucleic acid markers of the disease state identified in the present invention.

In one broad aspect, the present invention comprises an isolated nucleic acid of a sequence comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:4, SEQ ID NO:2, SEQ ID NO:29, and the sequences identified in Genebank Accession #s D87451, T03013, X03558, M28130, and Y00787. The invention further broadly comprises an isolated nucleic acid of between 17 and 100 bases in length, either identical to or complementary with portions of the above mentioned isolated nucleic acids. Such isolated nucleic acids may themselves be used as probes for

human disease markers, or may be used to design probes and primers specific for disease state markers. The invention further broadly comprises an isolated nucleic acid of between 17 bases to the full sequence length, either identical to or complementary with portions of the above mentioned isolated nucleic acids.

In another broad aspect, the present invention comprises proteins and peptides with amino acid sequences encoded by the aforementioned isolated nucleic acids. The proteins and peptides may be directly detected in the practice of the invention, or used for antibody production.

The invention also broadly comprises methods for identifying biomarkers for use in prognostic or diagnostic assays of a disease state, using the technique of RNA fingerprinting to identify RNAs that are differentially expressed between individuals with the disease state *versus* normal individuals. In the practice of the method, one may use random hexamers, arbitrarily chosen oligonucleotides, promiscuous oligonucleotide primers or anchoring primers, as well as oligonucleotide primers specific for known gene sequences for the reverse transcription step and/or for the amplification step. The term "promiscuous oligonucleotide primers" as used herein denotes oligonucleotides that are statistically designed to sample sequence complexity in mRNAs, or open reading frames of mRNAs without bias as applied in a PCR based RNA fingerprinting technique. The use of promiscuous primers is preferred because such use increases the sampling rate of RNA for fingerprinting by increasing the displayed fingerprint complexity. This increases the rate at which differentially expressed mRNAs can be discovered. The use of promiscuous oligonucleotide primers as disclosed herein will be evident to one of skill in the art in light of the publication by Lopez-Nieto and Nigam, *Nature Biotechnology* 14:857-861, 1996, (incorporated in pertinent part herein by reference). In certain embodiments the terms "random hexamers" or "small random oligonucleotides" refer to primers of random or semi-random nucleotide sequence of about 6 bases in length, though in certain embodiments the length of the primers may be of any length previously described for "primers". In certain aspects of the invention "arbitrarily chosen oligonucleotides" may refer to primers that are selected at the discretion of one skilled in the art, and may be of random or nonrandom sequence. In certain other embodiments "arbitrarily chosen oligonucleotides" may refer to primers as described by Welsh *et al.*, 1992, incorporated herein by reference. Oligonucleotide sequences designed to bind to specific genes, IL-8 or PSA for example, may also be used in the practice of this method.

The present invention may be described in a broad aspect as a method for identifying serological markers for a human disease state. The method comprises the steps of providing human peripheral blood mRNAs; amplifying the mRNAs to provide nucleic acid amplification products; separating the nucleic acid amplification products; and identifying those mRNAs that are differentially expressed between normal individuals and individuals exhibiting a disease state. The described method may also comprise, in certain embodiments, the step of converting the RNAs into cDNAs using reverse transcriptase to detect and quantitate circulating cells induced by the disease state. In certain embodiments of the invention conversion of RNA into cDNAs using reverse transcriptase is referred to as a "reverse transcriptase" reaction. Methods of reverse transcribing RNA into cDNA are well known and described in Sambrook *et al.*, 1989. Alternative methods for reverse transcription utilize thermostable, RNA-dependent DNA polymerases. These methods are described in WO 90/07641, filed December 21, 1990, incorporated herein by reference. In certain other embodiments of the invention a "reverse transcriptase" reaction refers to additional steps of amplification of the RNA template or its cDNA product. Such step of amplification may include any methods known in the art of increasing the number of copies of RNA or DNA, as well as the methods described herein. Methods of amplification include the methods described in Davey *et al.*, EPA No. 329 822 (incorporated herein by reference in its entirely), as well as polymerase chain reaction or ligase chain reaction

The method described in the previous paragraph may be used to discover disease markers for any disease state that affects the peripheral blood lymphocytes. Such diseases include, but are not limited to metastatic or organ defined cancer, particularly metastatic prostate or breast cancer, asthma, lupus erythematosis, rheumatoid arthritis, multiple sclerosis, myasthenia gravis, autoimmune thyroiditis, amyotrophic lateral sclerosis (ALS or Lou Gehrig's disease), interstitial cystitis or prostatitis.

The invention further broadly comprises methods for detecting a disease state in biological samples, using nucleic acid amplification techniques with primers and hybridization probes selected to bind specifically to an isolated nucleic acid of a sequence comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:4, SEQ ID NO:2 and SEQ ID NO:29 and the sequences

identified in Genebank Accession #s D87451, T03013, X03558, M28130, and Y00787, thereby measuring the amounts of nucleic acid amplification products formed.

The invention further broadly comprises the prognosis and/or diagnosis of a disease state by measuring the amounts of nucleic acid amplification products formed. The amounts of nucleic amplification products identified in an individual patient may be compared with groups of normal individuals or individuals with an identified disease state. Diagnosis may be accomplished by finding that the patient's levels of disease state markers fall within the normal range, or within the range observed in individuals with the disease state. Further comparison with groups of individuals of varying disease state progression, such as metastatic vs. non-metastatic cancer, may provide a prognosis for the individual patient. The invention further broadly comprises kits for performing the above-mentioned procedures, containing amplification primers and/or hybridization probes.

The invention may be described therefore, in certain broad aspects as a method of detecting a human disease state, comprising the steps of detecting the quantity of a disease marker expressed in human peripheral blood and comparing the quantity of the said marker to the quantity expressed in peripheral blood of a normal individual, where a difference in quantity of expression is indicative of a disease state. In the practice of the method the disease marker may preferably be an mRNA, or even an mRNA amplified by an RNA polymerase reaction, for example. The mRNA may also be amplified by any other means such as reverse transcriptase polymerase chain reaction or the ligase chain reaction. The RNA may be detected by any means known in the art, such as by RNA fingerprinting, branched DNA or a nuclease protection assay, for example. Disease states that may be detected by the present method include any disease state for which a marker is known and may include metastatic cancer, particularly metastatic prostate cancer, asthma, lupus erythromatosis, rheumatoid arthritis, multiple sclerosis, myasthenia gravis, autoimmune thyroiditis, amyotrophic lateral sclerosis, interstitial cystitis or prostatitis.

In certain preferred embodiments of this method, the mRNA will comprise one or more of the sequences or the complements of the transcribed sequences disclosed herein as SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:4, SEQ ID NO:2, SEQ ID NO:29 and the sequences identified in Genebank Accession #s D87451, T03013, X03558, M28130, and Y00787, or the mRNA may comprise a product of the interleukin 8 (IL-8) gene.

The method of detecting a disease state described in the previous paragraphs may further comprise the steps of providing primers that selectively amplify the disease state marker, amplifying the nucleic acid with said primers to form nucleic acid amplification products, detecting the nucleic acid amplification products and measuring the amount of the nucleic acid amplification products formed. In the practice of certain embodiments of the method, the primers may be selected to specifically amplify a nucleic acid having a sequence comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:4, SEQ ID NO:2, SEQ ID NO:29 and the sequences identified in Genebank Accession #s D87451, T03013, X03558, M28130, and Y00787. In certain alternate embodiments, the marker may be a polypeptide, and may even be a polypeptide encoded by a nucleic acid sequence comprising a sequence disclosed herein as SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:4, SEQ ID NO:2, SEQ ID NO:29 and the sequences identified in Genebank Accession #s D87451, T03013, X03558, M28130, and Y00787, or it may be described in certain embodiments as a polypeptide encoded by the IL-8 gene. Detection of the disease state may be by detection of an antibody immunoreactive with said marker. It is also an embodiment of the invention that detection may be by a cellular bioassay, that responds to the presence of a biologically active agent such as IL-8, for example. In certain embodiments of the present invention a "bioassay" is any assay that measures or detects the presence of a compound or effector, such as a protein, polypeptide, or peptide product of an expressed marker gene, by its affect on a cell, organism, or biologically derived reagent or detection system. Bioassays that may be used in the present invention, include, but are not limited to, those described in Schroder *et al.*, 1990 and Yoshimura *et al.*, 1989, Kurdowska et al., 1997, Hedges et al., 1996, (all incorporated herein by reference), and all bioassays known in the art that can be used to detect the expressed markers.

The present invention broadly comprises production of antibodies specific for proteins or peptides encoded by SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:4, SEQ ID NO:2, SEQ ID NO:29 and the sequences identified in Genebank Accession #s D87451, T03013, X03558, M28130, and Y00787, and the use of those antibodies for diagnostic applications in detecting and diagnosing the disease state. The levels of such proteins present in the peripheral blood of a patient may be quantitated by conventional methods. Correlation of protein levels with the presence of a

human disease or the progression of a human disease may be accomplished as described above for nucleic acid markers of human disease.

Another broad aspect of the present invention comprises the detection and diagnosis of disease states, including BPH and prostate cancer, by combining measurement of levels of two or more disease state markers. A broad embodiment of the invention comprises combining measurement of serum IL-8 gene product with other markers of prostate disease, such as PSA, PAP, HK2, PSP_{94} and PSMA. Yet another broad aspect of the present invention comprises kits for detection and measurement of the levels of two or more disease state markers in biological samples. The skilled practitioner will realize that such kits may incorporate a variety of methodologies for detection and measurement of disease state markers, including but not limited to oligonucleotide probes, primers for nucleic acid amplification, antibodies which bind specifically to protein products of disease state marker genes, and other proteins or peptides which bind specifically to disease state marker gene products.

BRIEF DESCRIPTION OF THE DRAWINGS

The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

FIG. 1A. Relative quantitative RT-PCR of UC Bands #325-1 (intron 3-) and 325-2 (intron 3+) shows that the normally spliced form of IL-8 mRNA (intron 3-) is abundantly expressed in individuals with metastatic prostate cancer (M) compared with normal individuals (N). The amplification reactions were sampled at different cycle numbers. The alternatively spliced form of IL-8 mRNA (intron 3+) is more abundant in normal individuals than in patients with metastatic cancer. The data were normalized against ß-actin mRNA.

FIG. 1B. Relative quantitative RT-PCR of UC Bands #325-1 (intron 3-) and 325-2 (intron 3+) shows that the normally spliced form of IL-8 mRNA (intron 3-) is abundantly

expressed in individuals with metastatic prostate cancer (M) compared with a pool of normal individuals (N). The alternatively spliced form of IL-8 mRNA (intron 3+) is more abundant in normal individuals than in patients with metastatic cancer. The data were normalized against ß-actin mRNA.

FIG. 2. Ability of total PSA (t-PSA) to distinguish BPH and Stages A, B, & C prostate cancer.

FIG. 3. Ability of corrected free/total PSA (f/t PSA) ratio to distinguish BPH and Stages A, B, & C prostate cancer.

FIG. 4. Ability of UC325 (IL-8) to distinguish BPH and Stages A, B, & C prostate cancer.

FIG. 5. Ability of UC325 (IL-8) and t-PSA combined to distinguish BPH and Stages A, B, & C prostate cancer.

FIG. 6. Ability of UC325 (IL-8) and the f/t PSA ratio combined to distinguish BPH and Stages A, B, & C prostate cancer.

FIG. 7. Relative quantitative RT-PCR™ showing that UC331 mRNA is roughly seven times more abundant in the peripheral blood of individuals with recurrent metastatic breast or prostate cancer compared to UC331 mRNA levels from healthy volunteers. PCR™ amplification of a UC331 specific cDNA fragment was performed using the same pools of ß-actin normalized cDNAs as templates. PCR™ reactions were terminated after either 25, 28 or 31 cycles. Pools of cDNAs were constructed from peripheral blood RNAs from eight healthy volunteers (N), ten individuals with recurrent metastatic prostate cancer (P), or ten individuals with recurrent metastatic breast cancer (B). The intensity of the bands are proportional to the relative amounts of UC331 mRNA in the individuals from which these cDNA pools were constructed.

FIG. 8. PCR™ amplification of a UC332 specific cDNA fragment using the same pools of normalized cDNAs as templates. PCR™ reactions were terminated after either 25, 28 or 31 cycles. Pools of cDNAs were constructed from peripheral blood RNAs from eight healthy volunteers (N), ten individuals with recurrent metastatic prostate cancer (P), or ten individuals with recurrent metastatic breast cancer (B). The intensity of the bands are proportional to the relative amounts of UC332 mRNA in the individuals from which these cDNA pools were constructed.

DETAILED DESCRIPTION OF THE INVENTION

Terms used:

HK2:	human kallekrein 2 gene product
PAP:	prostatic acid phosphatase
PSA:	prostate specific antigen
PSMA:	prostate specific membrane antigen (Folic Acid Hydrolase)
PSP_{94}:	prostate secreted protein (94 kDa)
t-PSA:	total PSA
f/t (Free/Total PSA):	ratio of free to total PSA, measured in serum specimens with moderately elevated t-PSA
IL-8:	Interleukin-8 (UC 325)

SENSITIVITY = (True Positives/(True Positives + False Negatives); plotted on *y-axis* of ROC curve.

SPECIFICITY = (True Negatives)/(True Negatives + False Positives); plotted on *x-axis (as 1-Specificity)* of ROC curve

ROC:	Receiver Operator Character Curve; a means of plotting sensitivity and specificity over a range of cut-off (threshold) values.
BPH:	benign prostate hyperplasia (or hypertrophy)
CaP:	adenocarcinoma of the prostate
Stage A CaP:	*organ-confined* clinical stage of prostate cancer in which tumor is *not* palpable by a digital rectal exam (DRE) (Walsh & Worthington, 1995).

Stage B CaP: *organ-confined* clinical stage of prostate cancer in which tumor is palpable by a digital rectal exam and involves one or both lobes of the gland (Walsh & Worthington, 1995).

Stage C CaP: *non-organ-confined* clinical stage of prostate cancer in which tumor is palpable by a DRE and invades beyond the capsule and/or the seminal vesicles (Walsh & Worthington, 1995).

Stage D CaP: *non-organ-confined* clinical stage of prostate cancer characterized by metastases to lymph nodes, bone or other distant organ site (Walsh & Worthington, 1995).

The present invention concerns the early detection, diagnosis, and prognosis of human disease states. Markers of a disease state, in the form of isolated nucleic acids of specified sequences from the peripheral blood of individuals with the disease state, are disclosed. These markers are indicators of the disease state and are diagnostic for the presence of the disease state in patients. Such markers provide considerable advantages over the prior art in this field. Since they are detected in peripheral blood samples, it is not necessary to suspect that an individual exhibits the disease state before a sample may be taken. The detection methods disclosed are thus suitable for widespread screening of asymptomatic individuals. Further, the methods provide for sensitive detection of disease state markers that is relatively unaffected by the presence of normal, non-diseased cells in a biological sample such as peripheral blood.

It will be apparent that the nucleic acid sequences disclosed will find utility in a variety of applications in disease state detection, diagnosis, prognosis and treatment. Examples of such applications within the scope of the present disclosure comprise amplification of markers of the disease state using specific primers, detection of markers of the disease state by hybridization with oligonucleotide probes, incorporation of isolated nucleic acids into vectors, expression of vector-incorporated nucleic acids as RNA and protein, and development of immunologic reagents corresponding to marker encoded products.

It is important to note that UC-325 (IL-8) serology in combination with PSA and f/t PSA can more accurately differentially diagnose prostate cancer and BPH. This method provides

significant advantages over previous methodologies for detecting prostatic cancer, which often failed to differentiate between prostatic cancer and BPH.

A. Nucleic Acids

As described in Examples 1 through 4, the present disclosure provides five markers of a disease state, identified by RNA fingerprinting. These include two previously previously uncharacterized gene products, as well as nucleic acid products of the IL-8 (interleukin 8) and human elongation factor 1-alpha genes.

In one embodiment, the sequences of isolated nucleic acids disclosed herein find utility as hybridization probes or amplification primers. These nucleic acids may be used, for example, in diagnostic evaluation of tissue samples or employed to clone full length cDNAs or genomic clones corresponding thereto. In certain embodiments, these probes and primers comprise oligonucleotide fragments. Such fragments are of sufficient length to provide specific hybridization to an RNA or DNA sample extracted from tissue. The sequences typically will be 10-20 nucleotides, but may be longer. Longer sequences, *e.g.*, 40, 50, 100, 500 and even up to full length, are preferred for certain embodiments.

Nucleic acid molecules having contiguous stretches of about 10, 15, 17, 20, 30, 40, 50, 60, 75 or 100 or 500 nucleotides of a sequence comprising Genebank Accession numbers D87451, T03013, X03558, M28130, Y00787, SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:29 are contemplated. Molecules that are complementary to the above mentioned sequences and that bind to these sequences under high stringency conditions are also contemplated. These probes are useful in a variety of hybridization embodiments, such as Southern and northern blotting. In some cases, it is contemplated that probes may be used that hybridize to multiple target sequences without compromising their ability to effectively diagnose the disease state.

Various probes and primers may be designed around the disclosed nucleotide sequences. Primers may be of any length but, typically, are 10-20 bases in length. By assigning numeric values to a sequence, for example, the first residue is 1, the second residue is 2, *etc.*, an algorithm defining all primers may be proposed:

n to n + y

where n is an integer from 1 to the last number of the sequence and y is the length of the primer minus one (9 to 19), where n + y does not exceed the last number of the sequence. Thus, for a 10-mer, the probes correspond to bases 1 to 10, 2 to 11, 3 to 12 ... and so on. For a 15-mer, the probes correspond to bases 1 to 15, 2 to 16, 3 to 17 ... and so on. For a 20-mer, the probes correspond to bases 1 to 20, 2 to 21, 3 to 22 ... and so on.

The values of n in the algorithm above for each of the nucleic acid sequences is: SEQ ID NO:1, n= 253; SEQ ID NO:2, n= 183; SEQ ID NO:3, n= 387; SEQ ID NO:4, n= 366; SEQ ID NO:5, n= 598.

In certain embodiments, it is contemplated that multiple probes may be used for hybridization to a single sample. For example, an alternatively spliced form of IL-8 mRNA, containing intron 3, may be detected by probing human tissue samples with oligonucleotides specific for intron 3 and for exon portions of the IL-8 transcript. Hybridization with the intron 3 and exon sequences probe would be indicative of a normal individual and binding to only the exon probe would be indicative of metastatic prostate cancer.

The use of a hybridization probe of between 17 and 100 nucleotides in length allows the formation of a duplex molecule that is both stable and selective. Molecules having complementary sequences over stretches greater than 20 bases in length are generally preferred in order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of hybrid molecules. It is generally preferred to design nucleic acid molecules having stretches of 20 to 30 nucleotides, or even longer. Such fragments may be readily prepared by, for example, directly synthesizing the fragment by chemical means or by introducing selected sequences into recombinant vectors for recombinant production.

The complement of a nucleic acid sequence is well known in the art and is based on the anti-parallel, Watson-Crick pairing of nucleotides (bases) for a given nucleic acid polymer (strand). Two complementary strands of DNA are formed into a duplex by pairing of bases, e.g. "G" to "C", "C" to "G", "A" to "T" (in the case of DNA) or "U" (in the case of RNA) and all "T" or "U" to "A", in reverse 5' to 3' orientation (anti-parallel). As used herein therefore, the term "complement" defines a second strand of nucleic acid which will hybridize to a first strand of nucleic acid to form a duplex molecule in which base pairs are matched as G:C, C:G, A:T/U or T/U:A.

A complement may also be described as a fragment of DNA (nucleic acid segment) or a synthesized single stranded oligomer that may contain small mismatches or gaps when hybridized to its complement, but that is able to hybridize to the complementary DNA under high stringency conditions. To hybridize is understood to mean the forming of a double stranded molecule or a molecule with partial double stranded nature. High stringency conditions are those that allow hybridization between two homologous nucleic acid sequences, but precludes hybridization of random sequences. For example, hybridization at low temperature and/or high ionic strength is termed low stringency. Hybridization at high temperature and/or low ionic strength is termed high stringency. Low stringency is generally performed at 0.15 M to 0.9 M NaCl at a temperature range of 20°C to 50°C. High stringency is generally performed at 0.02 M to 0.15 M NaCl at a temperature range of 50°C to 70°C. It is understood that the temperature and ionic strength of a desired stringency are determined in part by the length of the particular probe, the length and base content of the target sequences, and to the presence of formamide, tetramethylammonium chloride or other solvents in the hybridization mixture. It is also understood that these ranges are mentioned by way of example only, and that the desired stringency for a particular hybridization reaction is often determined empirically by comparison to positive and negative controls.

Accordingly, the nucleotide sequences of the disclosure may be used for their ability to selectively form duplex molecules with complementary stretches of genes or RNAs or to provide primers for amplification of DNA or RNA from tissues. Depending on the application envisioned, it is preferred to employ varying conditions of hybridization to achieve varying degrees of selectivity of probe towards target sequence.

For applications requiring high selectivity, it is preferred to employ relatively stringent conditions to form the hybrids. For example, relatively low salt and/or high temperature conditions, such as provided by about 0.02 M to about 0.10 M NaCl at temperatures of about 50°C to about 70°C. Such high stringency conditions tolerate little, if any, mismatch between the probe and the template or target strand, and would be particularly suitable for isolating specific genes or detecting specific mRNA transcripts. It is generally appreciated that conditions may be rendered more stringent by the addition of increasing amounts of formamide.

For certain applications, for example, substitution of amino acids by site-directed mutagenesis, it is appreciated that lower stringency conditions are required. Under these conditions, hybridization may occur even though the sequences of probe and target strand are not perfectly complementary, but are mismatched at one or more positions. Conditions may be rendered less stringent by increasing salt concentration and decreasing temperature. For example, a medium stringency condition may be provided by about 0.1 to 0.25 M NaCl at temperatures of about 37°C to about 55°C, while a low stringency condition may be provided by about 0.15 M to about 0.9 M salt, at temperatures ranging from about 20°C to about 55°C. Thus, hybridization conditions may be readily manipulated depending on the desired results.

The following codon chart may be used, in a site-directed mutagenic scheme, to produce nucleic acids encoding the same or slightly different amino acid sequences of a given nucleic acid:

TABLE 1

Amino Acids			Codons					
Alanine	Ala	A	GCA	GCC	GCG	GCU		
Cysteine	Cys	C	UGC	UGU				
Aspartic acid	Asp	D	GAC	GAU				
Glutamic acid	Glu	E	GAA	GAG				
Phenylalanine	Phe	F	UUC	UUU				
Glycine	Gly	G	GGA	GGC	GGG	GGU		
Histidine	His	H	CAC	CAU				
Isoleucine	Ile	I	AUA	AUC	AUU			
Lysine	Lys	K	AAA	AAG				
Leucine	Leu	L	UUA	UUG	CUA	CUC	CUG	CUU
Methionine	Met	M	AUG					
Asparagine	Asn	N	AAC	AAU				
Proline	Pro	P	CCA	CCC	CCG	CCU		
Glutamine	Gln	Q	CAA	CAG				
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGU
Serine	Ser	S	AGC	AGU	UCA	UCC	UCG	UCU
Threonine	Thr	T	ACA	ACC	ACG	ACU		
Valine	Val	V	GUA	GUC	GUG	GUU		
Tryptophan	Trp	W	UGG					
Tyrosine	Tyr	Y	UAC	UAU				

In other embodiments, hybridization may be achieved under conditions of, for example, 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 10 mM dithiothreitol, at temperatures between approximately 20°C to about 37°C. Other hybridization conditions utilized may include approximately 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, at temperatures ranging from approximately 40°C to about 72°C.

In certain embodiments, it is preferred to employ isolated nucleic acids of the present disclosure in combination with an appropriate means, such as a label, for determining hybridization. A wide variety of appropriate indicator means are known in the art, including fluorescent, radioactive, enzymatic or other ligands, such as avidin/biotin, which are capable of being detected. In preferred embodiments, one may employ a fluorescent label or an enzyme tag such as urease, alkaline phosphatase or peroxidase, instead of radioactive or other environmentally undesirable reagents. In the case of enzyme tags, colorimetric indicator substrates are known which may be employed to provide a detection means visible to the human eye or spectrophotometrically, to identify specific hybridization with complementary nucleic acid-containing samples.

In general, it is contemplated that the hybridization probes described herein are useful both as reagents in solution hybridization, as in PCR, for detection of expression of corresponding genes, as well as in embodiments employing a solid phase. In embodiments involving a solid phase, the test DNA (or RNA) is adsorbed or otherwise affixed to a selected matrix or surface. This fixed, single-stranded nucleic acid is then subjected to hybridization with selected probes under selected conditions. The selected conditions depend on the particular circumstances based on the particular criteria required (depending, for example, on the G+C content, type of target nucleic acid, source of nucleic acid, size of hybridization probe, *etc.*). Following washing of the hybridized surface to remove non-specifically bound probe molecules, hybridization is detected, or even quantified, by means of the label.

It is understood that this disclosure is not limited to the particular probes disclosed herein and particularly is intended to encompass at least isolated nucleic acids that are hybridizable to nucleic acids comprising the disclosed sequences or that are functional sequence analogs of these nucleic acids. For example, a nucleic acid of partial sequence may be used to identify a structurally-related gene or the full length genomic or cDNA clone from which it is derived. Methods for generating cDNA and genomic libraries which may be used as a target for the above-described probes are known in the art (Sambrook *et al.*, 1989).

For applications in which the nucleic acid segments of the present disclosure are incorporated into vectors, such as plasmids, cosmids or viruses, these segments may be combined with other DNA sequences, such as promoters, polyadenylation signals, restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary

considerably. It is contemplated that a nucleic acid fragment of almost any length may be employed, with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol.

DNA segments encoding a specific gene may be introduced into recombinant host cells for expressing a specific structural or regulatory protein. Alternatively, through the application of genetic engineering techniques, subportions or derivatives of selected genes may be employed. Upstream regions containing regulatory regions such as promoter regions may be isolated and subsequently employed for expression of the selected gene.

Where an expression product is to be generated, it is possible for the nucleic acid sequence to be varied while retaining the ability to encode the same product. Reference to the codon chart, provided in Table 1, enables the design of any nucleic acid encoding the same protein or peptide product.

B. Encoded Proteins

Once the entire coding sequence of a marker-associated gene has been determined, the gene may be inserted into an appropriate expression system. The gene may be expressed in any number of different recombinant DNA expression systems to generate large amounts of the polypeptide product, which may then be purified and used to vaccinate animals to generate antisera which may also be useful in the practice of the disclosed invention. For example, polyclonal or monoclonal antibodies may be prepared that specifically bind to the protein product(s) of the marker-associated gene. Such antibodies may be incorporated into kits that may in turn be used for detection and diagnosis of the disease state in peripheral blood or other tissue samples.

Examples of expression systems known in the art include bacteria such as *E. coli*, yeast such as *Saccharomyces cerevisia* and *Pichia pastoris*, baculovirus, and mammalian expression systems such as in Cos or CHO cells. In one embodiment, polypeptides are expressed in *E. coli* and in baculovirus expression systems. A complete gene may be expressed or, alternatively, fragments of the gene encoding portions of polypeptide may be produced.

In one embodiment, the gene sequence encoding the polypeptide is analyzed to detect putative transmembrane sequences. Such sequences are typically very hydrophobic and are readily detected by the use of sequence analysis software, such as Lasergene (DNAstar, Madison, WI).

The presence of transmembrane sequences is often deleterious when a recombinant protein is synthesized in many expression systems, especially *E. coli*, as it leads to the production of insoluble aggregates that are difficult to renature into the native conformation of the protein. Deletion of transmembrane sequences typically does not significantly alter the conformation of the remaining protein structure.

Moreover, transmembrane sequences, being by definition embedded within a membrane, are inaccessible. Antibodies to these sequences will not prove useful for *in vivo* or *in situ* studies. Deletion of transmembrane-encoding sequences from the genes used for expression may be achieved by conventional techniques. For example, restriction enzyme sites may be used to excise the desired gene fragment, or PCR-type amplification may be used to amplify only the desired part of the gene.

In another embodiment, computer sequence analysis is used to determine the location of predicted major antigenic determinant epitopes of the polypeptide. Software capable of carrying out this analysis is readily available commercially. Such software typically uses conventional algorithms such as the Kyte/Doolittle or Hopp/Woods methods for locating hydrophilic sequences which are characteristically found on the surface of proteins and are, therefore, likely to act as antigenic determinants.

Once this analysis is made, polypeptides may be prepared which contain at least the essential features of the antigenic determinant and which may be employed in the generation of antisera against the polypeptide. Minigenes or gene fusions encoding these determinants may be constructed and inserted into expression vectors by conventional methods, for example, using PCR cloning methodology.

A gene or gene fragment encoding a polypeptide may be inserted into an expression vector by conventional subcloning techniques. In one embodiment, an *E. coli* expression vector is used which produces the recombinant polypeptide as a fusion protein, allowing rapid affinity purification of the protein. Examples of such fusion protein expression systems are the glutathione S-transferase system (Pharmacia, Piscataway, NJ), the maltose binding protein system (NEB, Beverley, MA), the FLAG system (IBI, New Haven, CT), and the 6xHis system (Qiagen, Chatsworth, CA).

Some of these systems produce recombinant polypeptides bearing only a small number of additional amino acids, which are unlikely to affect the antigenic character of the recombinant polypeptide. For example, both the FLAG system and the 6xHis system add only short sequences, both of which are known to be poorly antigenic and which do not adversely affect folding of the polypeptide to its native conformation. Other fusion systems produce polypeptide where it is desirable to excise the fusion partner from the desired polypeptide. In one embodiment, the fusion partner is linked to the recombinant polypeptide by a peptide sequence containing a specific recognition sequence for a protease. Examples of suitable sequences are those recognized by the Tobacco Etch Virus protease (Life Technologies, Gaithersburg, MD) or Factor Xa (New England Biolabs, Beverley, MA).

In another embodiment, the expression system used is one driven by the baculovirus polyhedron promoter. The gene encoding the polypeptide may be manipulated by conventional techniques in order to facilitate cloning into the baculovirus vector. One baculovirus vector is the pBlueBac vector (Invitrogen, Sorrento, CA). The vector carrying the gene for the polypeptide is transfected into *Spodoptera frugiperda* (Sf9) cells by conventional protocols, and the cells are cultured and processed to produce the recombinant antigen. See Summers *et al.*, A MANUAL OF METHODS FOR BACULOVIRUS VECTORS AND INSECT CELL CULTURE PROCEDURES, Texas Agricultural Experimental Station; U.S. Patent No. 4,215,051 (incorporated by reference).

As an alternative to recombinant polypeptides, synthetic peptides corresponding to the antigenic determinants may be prepared. Such peptides are at least six amino acid residues long, and may contain up to approximately 50 residues, which is the approximate upper length limit of automated peptide synthesis machines, such as those available from Applied Biosystems (Foster City, CA). Use of such small peptides for vaccination typically requires conjugation of the peptide to an immunogenic carrier protein such as hepatitis B surface antigen, keyhole limpet hemocyanin or bovine serum albumin. Methods for performing this conjugation are well known in the art.

In one embodiment, amino acid sequence variants of the polypeptide may be prepared. These may, for instance, be minor sequence variants of the polypeptide which arise due to natural variation within the population or they may be homologues found in other species. They also may be sequences which do not occur naturally but which are sufficiently similar that they function

similarly and/or elicit an immune response that cross-reacts with natural forms of the polypeptide. Sequence variants may be prepared by conventional methods of site-directed mutagenesis such as those described above for removing the transmembrane sequence.

Amino acid sequence variants of the polypeptide may be substitutional, insertional or deletion variants. Deletion variants lack one or more residues of the native protein which are not essential for function or immunogenic activity, and are exemplified by the variants lacking a transmembrane sequence described above. Another common type of deletion variant is one lacking secretory signal sequences or signal sequences directing a protein to bind to a particular part of a cell. An example of the latter sequence is the SH2 domain, which induces protein binding to phosphotyrosine residues.

Substitutional variants typically exchange one amino acid for another at one or more sites within the protein and may be designed to modulate one or more properties of the polypeptide, such as stability against proteolytic cleavage. Substitutions preferably are conservative, that is, one amino acid is replaced with another of similar shape and charge. Conservative substitutions are well known in the art and include, for example, the changes of: alanine to serine; arginine to lysine; asparagine to glutamine or histidine; aspartate to glutamate; cysteine to serine; glutamine to asparagine; glutamate to aspartate; histidine to asparagine or glutamine; isoleucine to leucine or valine; leucine to valine or isoleucine; lysine to arginine or glutamine; methionine to leucine or isoleucine; phenylalanine to tyrosine, leucine or methionine; serine to threonine; threonine to serine; tryptophan to tyrosine; tyrosine to tryptophan or phenylalanine; and valine to isoleucine or leucine.

Insertional variants include fusion proteins such as those used to allow rapid purification of the polypeptide and also may include hybrid proteins containing sequences from other homologous proteins and polypeptides. For example, an insertional variant may include portions of the amino acid sequence of the polypeptide from one species, together with portions of the homologous polypeptide from another species. Other insertional variants may include those in which additional amino acids are introduced within the coding sequence of the polypeptide. These typically are smaller insertions than the fusion proteins described above and are introduced, for example, to disrupt a protease cleavage site.

In one embodiment, major antigenic determinants of the polypeptide are identified by an empirical approach in which portions of the gene encoding the polypeptide are expressed in a recombinant host, and the resulting proteins tested for their ability to elicit an immune response. For example, PCR may be used to prepare a range of peptides lacking successively longer fragments of the C-terminus of the protein. The immunoprotective activity of each of these peptides then identifies those fragments or domains of the polypeptide which are essential for this activity. Further studies in which only a small number of amino acids are removed at each iteration then enables the location of the antigenic determinants of the polypeptide.

Another embodiment for the preparation of polypeptides according to the disclosure is the use of peptide mimetics. Mimetics are peptide-containing molecules which mimic elements of protein secondary structure. See, for example, Johnson *et al.*, "Peptide Turn Mimetics" in *BIOTECHNOLOGY AND PHARMACY*, Pezzuto *et al.*, Eds., Chapman and Hall, New York (1993). The underlying rationale behind the use of peptide mimetics is that the peptide backbone of proteins exists chiefly to orient amino acid side chains in such a way as to facilitate molecular interactions, such as those of antibody and antigen. A peptide mimetic is expected to permit molecular interactions similar to the natural molecule.

Successful applications of the peptide mimetic concept have thus far focused on mimetics of β-turns within proteins, which are known to be highly antigenic. Likely β-turn structure within an polypeptide may be predicted by computer-based algorithms as discussed above. Once the component amino acids of the turn are determined, peptide mimetics may be constructed to achieve a similar spatial orientation of the essential elements of the amino acid side chains.

C. Preparation of Antibodies Specific for Encoded Proteins

1. Expression of Proteins from Cloned cDNAs

The cDNAs of sequences comprising Genebank Accession numbers D87451, T03013, X03558, M28130, Y00787, SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:29 may be expressed as encoded peptides or proteins. The engineering of DNA segment(s) for expression in a prokaryotic or eukaryotic system may be performed by techniques generally known in the art of recombinant expression. It is believed that

virtually any expression system may be employed in the expression of the claimed isolated nucleic acids.

Both cDNA and genomic sequences are suitable for eukaryotic expression, as the host cell generally processes the genomic transcripts to yield functional mRNA for translation into protein. In addition, it is possible to use partial sequences for generation of antibodies against discrete portions of a gene product, even when the entire sequence of that gene product remains unknown. Computer programs are available to aid in the selection of regions which have potential immunologic significance. Software capable of carrying out this analysis is readily available commercially, for example MacVector (IBI, New Haven, CT). The software typically uses conventional algorithms such as the Kyte/Doolittle or Hopp/Woods methods for locating hydrophilic sequences which are characteristically found on the surface of proteins and are therefore likely to act as antigenic determinants.

It may be more convenient to employ as the recombinant gene a cDNA version of the gene. It is believed that the use of a cDNA version provides advantages in that the size of the gene is generally much smaller and more readily employed to transfect the targeted cell than a genomic gene, which is typically up to an order of magnitude larger than the cDNA gene. However, the possibility of employing a genomic version of a particular gene or fragments thereof is specifically contemplated.

As used herein, the terms "engineered" and "recombinant" cells are intended to refer to a cell into which an exogenous DNA segment or gene, such as a cDNA or gene has been introduced. Therefore, engineered cells are distinguishable from naturally occurring cells which do not contain a recombinantly introduced exogenous DNA segment or gene. Engineered cells are thus cells having a gene or genes introduced through the hand of man. Recombinant cells include those having an introduced cDNA or genomic gene, and also include genes positioned adjacent to a promoter not naturally associated with the particular introduced gene.

To express a recombinant encoded protein or peptide, whether mutant or wild-type, in accordance with the present disclosure one prepares an expression vector that comprises one of the claimed isolated nucleic acids under the control of, or operatively linked to, one or more promoters. To bring a coding sequence "under the control of" a promoter, or to "operatively link" to a promoter, one positions the 5' end of the transcription initiation site of the transcriptional reading

frame generally between about 1 and about 50 nucleotides "downstream" of (*i.e.*, 3' of) the chosen promoter. The "upstream" promoter stimulates transcription of the DNA and promotes expression of the encoded recombinant protein. This is the meaning of "recombinant expression" in this context.

Many conventional techniques are available to construct expression vectors containing the appropriate nucleic acids and transcriptional/translational control sequences in order to achieve protein or peptide expression in a variety of host-expression systems. Cell types available for expression include, but are not limited to, bacteria, such as *E. coli* and *B. subtilis* transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors.

Certain examples of prokaryotic hosts are *E. coli* strain RR1, *E. coli* LE392, *E. coli* B, *E. coli* X 1776 (ATCC No. 31537) as well as *E. coli* W3110 (F-, lambda-, prototrophic, ATCC No. 273325); bacilli such as *Bacillus subtilis*; and other enterobacteriaceae such as *Salmonella typhimurium*, *Serratia marcescens*, and various *Pseudomonas* species.

In general, plasmid vectors containing replicon and control sequences which are derived from species compatible with the host cell are used in connection with these hosts. The vector ordinarily carries a replication site, as well as marking sequences which are capable of providing phenotypic selection in transformed cells. For example, *E. coli* is often transformed using pBR322, a plasmid derived from an *E. coli* species. pBR322 contains genes for ampicillin and tetracycline resistance and thus provides easy means for identifying transformed cells. The pBR plasmid, or other microbial plasmid or phage must also contain, or be modified to contain, promoters which may be used by the microbial organism for expression of its own proteins.

In addition, phage vectors containing replicon and control sequences that are compatible with the host microorganism may be used as transforming vectors in connection with these hosts. For example, the phage lambda GEM™-11 may be utilized in making a recombinant phage vector which may be used to transform host cells, such as *E. coli* LE392.

Further useful vectors include pIN vectors (Inouye *et al.*, 1985); and pGEX vectors, for use in generating glutathione S-transferase (GST) soluble fusion proteins for later purification and separation or cleavage. Other suitable fusion proteins are those with ß-galactosidase, ubiquitin, or the like.

Promoters that are most commonly used in recombinant DNA construction include the β-lactamase (penicillinase), lactose and tryptophan (trp) promoter systems. While these are the most commonly used, other microbial promoters have been discovered and utilized, and details concerning their nucleotide sequences have been published, enabling their ligation into plasmid vectors.

For expression in *Saccharomyces*, the plasmid YRp7, for example, is commonly used. This plasmid already contains the *trp*1 gene which provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, for example ATCC No. 44076 or PEP4-1. The presence of the *trp*1 lesion as a characteristic of the yeast host cell genome then provides an effective environment for detecting transformation by growth in the absence of tryptophan.

Suitable promoting sequences in yeast vectors include the promoters for 3-phosphoglycerate kinase or other glycolytic enzymes, such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase. In constructing suitable expression plasmids, the termination sequences associated with these genes are also ligated into the expression vector 3' of the sequence desired to be expressed to provide polyadenylation of the mRNA and termination.

Other suitable promoters, which have the additional advantage of transcription controlled by growth conditions, include the promoter region for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, and the aforementioned glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization.

In addition to micro-organisms, cultures of cells derived from multicellular organisms may also be used as hosts. In principle, any such cell culture is workable, whether from vertebrate or invertebrate culture. In addition to mammalian cells, these include insect cell systems infected with recombinant virus expression vectors (e.g., baculovirus); and plant cell systems infected with recombinant virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (e.g., Ti plasmid) containing one or more coding sequences.

In a useful insect system, *Autographia californica* nuclear polyhidrosis virus (AcNPV) is used as a vector to express foreign genes. The virus grows in *Spodoptera frugiperda* cells. The isolated nucleic acid coding sequences are cloned into non-essential regions (for example the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (for example the polyhedrin promoter). Successful insertion of the coding sequences results in the inactivation of the polyhedrin gene and production of non-occluded recombinant virus (*i.e.*, virus lacking the proteinaceous coat coded for by the polyhedrin gene). These recombinant viruses are then used to infect *Spodoptera frugiperda* cells in which the inserted gene is expressed (e.g., U.S. Patent No. 4,215,051 (Smith)).

Examples of useful mammalian host cell lines are VERO and HeLa cells, Chinese hamster ovary (CHO) cell lines, W138, BHK, COS-7, 293, HepG2, 3T3, RIN and MDCK cell lines. In addition, a host cell strain may be chosen that modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications (e.g., glycosylation) and processing (e.g., cleavage) of protein products may be important for the function of the encoded protein.

Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins. Appropriate cells lines or host systems may be chosen to help ensure the correct modification and processing of the foreign protein expressed. Expression vectors for use in mammalian cells ordinarily include an origin of replication, a promoter located in front of the gene to be expressed, along with any necessary ribosome binding sites, RNA splice sites, polyadenylation site, and transcriptional terminator sequences. The origin of replication may be provided either by construction of the vector to include an exogenous origin, such as may be derived from SV40 or other viral (*e.g.*, Polyoma, Adeno, VSV, BPV) source, or may be provided by the host cell chromosomal replication mechanism. If the vector is integrated into the host cell chromosome, the latter is often sufficient.

The promoters may be derived from the genome of mammalian cells (e.g., metallothionein promoter) or from mammalian viruses (e.g., the adenovirus late promoter; the vaccinia virus 7.5K promoter). Further, it is also possible to utilize promoter or control sequences normally associated with the gene sequence of interest, provided such control sequences are compatible with the host cell systems.

A number of viral based expression systems may be utilized. For example, commonly used promoters are derived from polyoma, Adenovirus 2, and most frequently Simian Virus 40 (SV40). The early and late promoters of SV40 virus are particularly useful because both are obtained easily from the virus as a fragment which also contains the SV40 viral origin of replication. Smaller or larger SV40 fragments may also be used, provided there is included the approximately 250 bp sequence extending from the Hind III site toward the Bgl I site located in the viral origin of replication.

In cases where an adenovirus is used as an expression vector, the coding sequences may be ligated to an adenovirus transcription/ translation control complex, e.g., the late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion in a non-essential region of the viral genome (e.g., region E1 or E3) results in a recombinant virus that is viable and capable of expressing proteins in infected hosts.

Specific initiation signals may also be required for efficient translation of the claimed isolated nucleic acid coding sequences. These signals include the ATG initiation codon and adjacent sequences. Exogenous translational control signals, including the ATG codon, may additionally need to be provided. This need is readily determinable and the necessary signals readily provided. It is well known that the initiation codon must be in-frame (or in-phase) with the reading frame of the desired coding sequence to help ensure translation of the entire insert. These exogenous translational control signals and initiation codons may be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements or transcription terminators (Bittner *et al.*, 1987).

In eukaryotic expression, it is typically preferred to incorporate into the transcriptional unit an appropriate polyadenylation site (e.g., 5'-AATAAA-3') if one was not contained within the original cloned segment. Typically, the poly A addition site is placed about 30 to 2000 nucleotides "downstream" of the termination site of the protein at a position prior to transcription termination.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines that stably express constructs encoding proteins may be engineered. Rather than using expression vectors that contain viral origins of replication, host cells may be transformed with vectors controlled by appropriate expression control elements (e.g.,

promoter or enhancer sequences, transcription terminators, polyadenylation sites, etc.), and a selectable marker. Following the introduction of foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched medium and then are switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the transformant and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci which in turn may be cloned and expanded into cell lines.

A number of selection systems may be used, including, but not limited, to the herpes simplex virus thymidine kinase (Wigler *et al.*, 1977), hypoxanthine-guanine phosphoribosyltransferase (Szybalska *et al.*, 1962) and adenine phosphoribosyltransferase genes (Lowy *et al.*, 1980), in tk-, hgprt- or aprt- cells, respectively. Also, antimetabolite resistance may be used as the basis of selection for dhfr, that confers resistance to methotrexate (Wigler *et al.*, 1980; O'Hare *et al.*, 1981); gpt, that confers resistance to mycophenolic acid (Mulligan *et al.*, 1981); neo, that confers resistance to the aminoglycoside G-418 (Colberre-Garapin *et al.*, 1981); and hygro, that confers resistance to hygromycin.

It is contemplated that the isolated nucleic acids of the disclosure may be "overexpressed", *i.e.*, expressed in increased levels relative to their natural expression in normal human cells, or even relative to the expression of other proteins in the recombinant host cell. Such overexpression may be assessed by a variety of methods, including radio-labeling and/or protein purification. However, simple and direct methods are preferred, for example, those involving SDS/PAGE and protein staining or Western blotting, followed by quantitative analyses, such as densitometric scanning of the resultant gel or blot. A specific increase in the level of the recombinant protein or peptide in comparison to the level in natural human cells is indicative of overexpression, as is a relative abundance of the specific protein in relation to the other proteins produced by the host cell and, e.g., visible on a gel.

2. *Purification of Expressed Proteins*

Further aspects of the present disclosure concern the purification, and in particular embodiments, the substantial purification, of an encoded protein or peptide. The term "purified protein or peptide " as used herein, is intended to refer to a composition, isolatable from other components, wherein the protein or peptide is purified to any degree relative to its naturally-

obtainable state, *i.e.*, in this case, relative to its purity within a cell extract. A purified protein or peptide therefore also refers to a protein or peptide, free from the environment in which it may naturally occur.

Generally, "purified" refers to a protein or peptide composition which has been subjected to fractionation to remove various other components, and which composition substantially retains its expressed biological activity. Where the term "substantially purified" is used, this refers to a composition in which the protein or peptide forms the major component of the composition, such as constituting about 50% or more of the proteins in the composition.

Various methods for quantifying the degree of purification of the protein or peptide are known in the art. These include, for example, determining the specific activity of an active fraction, or assessing the number of polypeptides within a fraction by SDS/PAGE analysis. A preferred method for assessing the purity of a fraction is to calculate the specific activity of the fraction, to compare it to the specific activity of the initial extract, and to thus calculate the degree of purity, assessed by a "-fold purification number". The actual units used to represent the amount of activity is dependent upon the particular assay technique chosen to follow the purification and whether or not the expressed protein or peptide exhibits an enzymatic or other activity.

Various techniques suitable for use in protein purification are known in the art. These include, for example, precipitation with ammonium sulfate, PEG, antibodies and the like or by heat denaturation, followed by centrifugation; chromatography steps such as ion exchange, gel filtration, reverse phase, hydroxylapatite and affinity chromatography; isoelectric focusing; gel electrophoresis; and combinations of such and other techniques. As is generally known in the art, it is believed that the order of conducting the various purification steps may be changed, or that certain steps may be omitted, and still result in a suitable method for the preparation of a substantially purified protein or peptide.

There is no general requirement that a protein or peptide always be provided in its most purified state. Indeed, it is contemplated that less substantially purified products have utility in certain embodiments. Partial purification may be accomplished by using fewer purification steps in combination, or by utilizing different forms of the same general purification scheme. For example, it is appreciated that a cation-exchange column chromatography performed utilizing an HPLC apparatus generally results in a greater -fold purification than the same technique utilizing a low

pressure chromatography system. Methods exhibiting a lower degree of relative purification may have advantages in total recovery of protein product, or in maintaining the activity of an expressed protein.

It is known that the migration of a polypeptide may vary, sometimes significantly, with different conditions of SDS/PAGE (Capaldi *et al.*, *Biochem. Biophys. Res. Comm.*, *76*:425, 1977). It is therefore appreciated that under differing electrophoresis conditions, the apparent molecular weights of purified or partially purified expression products may vary.

3. Antibody Generation

For some embodiments, it is preferred to produce antibodies that bind with high specificity to the protein product(s) of an isolated nucleic acid of a sequence comprising Genebank Accession numbers D87451, T03013, X03558, M28130, Y00787, SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:29. Means for preparing and characterizing antibodies are well known in the art (See, e.g., Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, 1988).

Methods for generating polyclonal antibodies are well known in the art. Briefly, a polyclonal antibody is prepared by immunizing an animal with an immunogenic composition and collecting antisera from that immunized animal. A wide range of animal species may be used for the production of antisera, including rabbits, mice, rats, hamsters, guinea pigs or goats. Because of the relatively large blood volume of rabbits, a rabbit is a preferred choice for production of polyclonal antibodies.

As is well known in the art, a given composition may vary in its immunogenicity. It is often necessary therefore to boost the host immune system, as may be achieved by coupling a peptide or polypeptide immunogen to a carrier. Exemplary and preferred carriers are keyhole limpet hemocyanin (KLH) and bovine serum albumin (BSA). Other albumins such as ovalbumin, mouse serum albumin or rabbit serum albumin may also be used as carriers. Means for conjugating a polypeptide to a carrier protein are well known in the art and include glutaraldehyde, m-maleimidobenzoyl-N-hydroxysuccinimideester, carbodiimide and bis-biazotized benzidine.

As is also well known in the art, the immunogenicity of a particular immunogen composition may be enhanced by the use of non-specific stimulators of the immune response,

known as adjuvants. Exemplary and preferred adjuvants include complete Freund's adjuvant (a non-specific stimulator of the immune response containing killed *Mycobacterium tuberculosis)*, incomplete Freund's adjuvants and aluminum hydroxide adjuvant.

The amount of immunogen composition used in the production of polyclonal antibodies varies with the nature of the immunogen as well as the animal used for immunization. A variety of routes may be used to administer the immunogen (subcutaneous, intramuscular, intradermal, intravenous and intraperitoneal). The production of polyclonal antibodies may be monitored by sampling blood of the immunized animal at various points following immunization. A second, booster, injection may also be given. The process of boosting and titering is repeated until a suitable titer is achieved. When a desired level of immunogenicity is obtained, the immunized animal may be bled and the serum isolated and stored, and/or the animal may be used to generate monoclonal antibodies. For production of rabbit polyclonal antibodies, the animal may be bled through an ear vein or alternatively by cardiac puncture. The removed blood is allowed to coagulate and then centrifuged to separate serum components from whole cells and blood clots. The serum may be used as is for various applications or else a particular antibody fraction may be purified by well-known methods, such as affinity chromatography using another antibody or a peptide bound to a solid matrix.

Monoclonal antibodies (MAbs) may be readily prepared through use of well-known techniques, such as those exemplified in U.S. Patent 4,196,265, incorporated herein by reference. Typically, this technique involves immunizing a suitable animal with a selected immunogen composition, e.g., a purified or partially purified expressed protein, polypeptide or peptide. The immunizing composition is administered in a manner effective to stimulate antibody producing cells, as described above.

The methods for generating monoclonal antibodies (MAbs) generally begin along the same lines as those for preparing polyclonal antibodies. Rodents such as mice and rats are preferred animals, however, the use of rabbit, sheep or frog cells is also possible. The use of rats may provide certain advantages (Goding, 1986, pp. 60-61), but mice are preferred, with the BALB/c mouse being most preferred as this generally gives a higher percentage of stable fusions.

The animals are injected with antigen as described above. The antigen may be coupled to carrier molecules such as keyhole limpet hemocyanin if necessary. The antigen is typically mixed

with adjuvant, such as Freund's complete or incomplete adjuvant. Booster injections with the same antigen typically occur at approximately two-week intervals.

Following immunization, somatic cells with the potential for producing antibodies, specifically B lymphocytes (B cells), are selected for use in the MAb generating protocol. These cells may be obtained from biopsied spleens, tonsils or lymph nodes, or from a peripheral blood sample. Spleen cells and peripheral blood cells are preferred, the former because they are a rich source of antibody-producing cells that are in the dividing plasmablast stage, and the latter because peripheral blood is easily accessible. Often, a panel of animals are immunized and the spleen of the animal with the highest antibody titer is removed and the spleen lymphocytes obtained by homogenizing the spleen with a syringe. Typically, a spleen from an immunized mouse contains approximately 5×10^7 to 2×10^8 lymphocytes.

The antibody-producing B lymphocytes from the immunized animal are then fused with cells of an immortal myeloma cell, generally one of the same species as the animal that was immunized. Myeloma cell lines suited for use in hybridoma-producing fusion procedures preferably are non-antibody-producing, have high fusion efficiency, and enzyme deficiencies that render then incapable of growing in selective media which support the growth of only the desired fused cells (hybridomas).

Any one of a number of myeloma cells may be used, as are known in the art (Goding, pp. 65-66, 1986; Campbell, pp. 75-83, 1984). For example, where the immunized animal is a mouse, one may use P3-X63/Ag8, X63-Ag8.653, NS1/1.Ag 4 1, Sp210-Ag14, FO, NSO/U, MPC-11, MPC11-X45-GTG 1.7 and S194/5XX0 Bul; for rats, one may use R210.RCY3, Y3-Ag 1.2.3, IR983F and 4B210; and U-266, GM1500-GRG2, LICR-LON-HMy2 and UC729-6 are all useful in connection with human cell fusions.

One preferred murine myeloma cell is the NS-1 myeloma cell line (also termed P3-NS-1-Ag4-1), which is readily available from the NIGMS Human Genetic Mutant Cell Repository by requesting cell line repository number GM3573. Another mouse myeloma cell line that may be used is the 8-azaguanine-resistant mouse murine myeloma SP2/0 non-producer cell line.

Methods for generating hybrids of antibody-producing spleen or lymph node cells and myeloma cells usually comprise mixing somatic cells with myeloma cells in a 2:1 proportion, though the proportion may vary from about 20:1 to about 1:1, respectively, in the presence of an

agent or agents (chemical or electrical) that promote the fusion of cell membranes. Fusion methods using Sendai virus have been described by Kohler and Milstein (1975; 1976), and those using polyethylene glycol (PEG), such as 37% (v/v) PEG, by Gefter *et al.* (1977). The use of electrically induced fusion methods is also appropriate (Goding pp. 71-74, 1986).

Fusion procedures usually produce viable hybrids at low frequencies, about 1×10^{-6} to 1×10^{-8}. However, this does not pose a problem, as the viable, fused hybrids are differentiated from the parental, unfused cells (particularly the unfused myeloma cells that would normally continue to divide indefinitely) by culturing in a selective medium. The selective medium is generally one that contains an agent that blocks the *de novo* synthesis of nucleotides in the tissue culture media. Exemplary and preferred agents are aminopterin, methotrexate, and azaserine. Aminopterin and methotrexate block *de novo* synthesis of both purines and pyrimidines, whereas azaserine blocks only purine synthesis. Where aminopterin or methotrexate is used, the media is supplemented with hypoxanthine and thymidine as a source of nucleotides (HAT medium). Where azaserine is used, the media is supplemented with hypoxanthine.

The preferred selection medium is HAT. Only cells capable of operating nucleotide salvage pathways are able to survive in HAT medium. The myeloma cells are defective in key enzymes of the salvage pathway, e.g., hypoxanthine phosphoribosyl transferase (HPRT), and they cannot survive. The B cells may operate this pathway, but they have a limited life span in culture and generally die within about two weeks. Therefore, the only cells that may survive in the selective media are those hybrids formed from myeloma and B cells.

This culturing provides a population of hybridomas from which specific hybridomas are selected. Typically, selection of hybridomas is performed by culturing the cells by single-clone dilution in microtiter plates, followed by testing the individual clonal supernatants (after about two to three weeks) for the desired reactivity. The assay should be sensitive, simple and rapid, such as radioimmunoassays, enzyme immunoassays, cytotoxicity assays, plaque assays, dot immunobinding assays, and the like.

The selected hybridomas are then serially diluted and cloned into individual antibody-producing cell lines, which clones may then be propagated indefinitely to provide MAbs. The cell lines may be exploited for MAb production in two basic ways. A sample of the hybridoma may be injected (often into the peritoneal cavity) into a histocompatible animal of the type that was

used to provide the somatic and myeloma cells for the original fusion. The injected animal develops tumors secreting the specific monoclonal antibody produced by the fused cell hybrid. The body fluids of the animal, such as serum or ascites fluid, may then be tapped to provide MAbs in high concentration. The individual cell lines may also be cultured *in vitro*, where the MAbs are naturally secreted into the culture medium from which they may be readily obtained in high concentrations. MAbs produced by either means may be further purified as needed, using filtration, centrifugation and various chromatographic methods such as HPLC or affinity chromatography.

Large amounts of the monoclonal antibodies of the present disclosure may also be obtained by multiplying hybridoma cells *in vivo*. Cell clones are injected into mammals which are histocompatible with the parent cells, *e.g.*, syngeneic mice, to cause growth of antibody-producing tumors. Optionally, the animals are primed with a hydrocarbon, especially oils such as pristane (tetramethylpentadecane)prior to injection.

In accordance with the present invention, fragments of monoclonal antibodies may be obtained by methods which include digestion of monoclonal antibodies with enzymes such as pepsin or papain and/or cleavage of disulfide bonds by chemical reduction. Alternatively, monoclonal antibody fragments encompassed by the present disclosure may be synthesized using an automated peptide synthesizer.

The monoclonal conjugates of the present disclosure are prepared by methods known in the art, *e.g.*, by reacting a monoclonal antibody prepared as described above with, for instance, an enzyme in the presence of a coupling agent such as glutaraldehyde or periodate. Conjugates with fluorescein markers are prepared in the presence of these coupling agents or by reaction with an isothiocyanate. Conjugates with metal chelates are similarly produced. Other moieties to which antibodies may be conjugated include radionuclides such as ^{3}H, ^{125}I, ^{131}I ^{32}P, ^{35}S, ^{14}C, ^{51}Cr, ^{36}Cl, ^{57}Co, ^{58}Co, ^{59}Fe, ^{75}Se, ^{152}Eu, and ^{99m}Tc, or other useful labels which may be conjugated to antibodies. Radioactively labeled monoclonal antibodies of the present disclosure are produced according to well-known methods in the art. For instance, monoclonal antibodies may be iodinated by contact with sodium or potassium iodide and a chemical oxidizing agent such as sodium hypochlorite, or an enzymatic oxidizing agent, such as lactoperoxidase. Monoclonal antibodies according to the disclosure may be labeled with technetium^{99} by ligand exchange process, for example, by reducing pertechnate with stannous solution, chelating the reduced technetium onto a

Sephadex column and applying the antibody to this column or by direct labeling techniques, *e.g.*, by incubating pertechnate, a reducing agent such as $SNCl_2$, a buffer solution such as sodium-potassium phthalate solution, and the antibody.

It will be appreciated that monoclonal or polyclonal antibodies specific for proteins that are preferentially expressed in the peripheral blood of individuals with the disease state have utilities in several types of applications. These may include the production of diagnostic kits for use in detecting or diagnosing the disease state. It will be recognized that such uses are within the scope of the present invention.

D. Immunodetection Assays

1. Immunodetection Methods

In still further embodiments, the present disclosure concerns immunodetection methods for binding, purifying, removing, quantifying or otherwise generally detecting biological components. The encoded proteins or peptides of the present disclosure may be employed to detect antibodies having reactivity therewith, or, alternatively, antibodies prepared in accordance with the present invention, may be employed to detect the encoded proteins or peptides. The steps of various useful immunodetection methods have been described in the scientific literature, such as, e.g., Nakamura *et al.* (1987).

In general, the immunobinding methods include obtaining a sample suspected of containing a protein, peptide or antibody, and contacting the sample with an antibody or protein or peptide in accordance with the present invention, as the case may be, under conditions effective to allow the formation of immunocomplexes.

The immunobinding methods include methods for detecting or quantifying the amount of a reactive component in a sample, which methods require the detection or quantitation of any immune complexes formed during the binding process. Here, one obtains a sample suspected of containing a disease state-marker encoded protein, peptide or a corresponding antibody, and contacts the sample with an antibody or encoded protein or peptide, as the case may be, and then detects or quantifies the amount of immune complex formed under the specific conditions.

In terms of antigen detection, the biological sample analyzed would ordinarily consist of peripheral blood. However, it may be any sample that is suspected of containing a disease state-

specific antigen, such as a lymph node tissue section or specimen, a homogenized tissue extract, an isolated cell, a cell membrane preparation, separated or purified forms of any of the above protein-containing compositions, or any other biological fluid that comes into contact with diseased tissues, including lymphatic fluid, urine and even seminal fluid.

Contacting the chosen biological sample with the protein, peptide or antibody under conditions effective and for a period of time sufficient to allow the formation of immune complexes (primary immune complexes) is generally a matter of simply adding the composition to the sample and incubating the mixture for a period of time long enough for the antibodies to form immune complexes with, *i.e.*, to bind to, any antigens present. After this time, the sample-antibody composition, such as a tissue section, ELISA plate, dot blot or Western blot, is generally washed to remove any non-specifically bound antibody species, allowing only those antibodies specifically bound within the primary immune complexes to be detected.

In general, the detection of immunocomplex formation is well known in the art and may be achieved through the application of numerous approaches. These methods are generally based upon the detection of a label or marker, such as any radioactive, fluorescent, biological or enzymatic tags or labels of conventional use in the art. U.S. Patents concerning the use of such labels include 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241, each incorporated herein by reference. Of course, one may find additional advantages through the use of a secondary binding ligand such as a second antibody or a biotin/avidin ligand binding arrangement, as is known in the art.

The encoded protein, peptide or corresponding antibody employed in the detection may itself be linked to a detectable label, wherein one would then simply detect this label, thereby allowing the amount of the primary immune complexes in the composition to be determined.

Alternatively, the first added component that becomes bound within the primary immune complexes may be detected by means of a second binding ligand that has binding affinity for the encoded protein, peptide or corresponding antibody. In these cases, the second binding ligand may be linked to a detectable label. The second binding ligand is itself often an antibody, which may thus be termed a "secondary" antibody. The primary immune complexes are contacted with the labeled, secondary binding ligand, or antibody, under conditions effective and for a period of time sufficient to allow the formation of secondary immune complexes. The secondary immune

complexes are then generally washed to remove any non-specifically bound labeled secondary antibodies or ligands, and the remaining label in the secondary immune complexes is then detected.

Further methods include the detection of primary immune complexes by a two step approach. A second binding ligand, such as an antibody, that has binding affinity for the encoded protein, peptide or corresponding antibody is used to form secondary immune complexes, as described above. After washing, the secondary immune complexes are contacted with a third binding ligand or antibody that has binding affinity for the second antibody, again under conditions effective and for a period of time sufficient to allow the formation of immune complexes (tertiary immune complexes). The third ligand or antibody is linked to a detectable label, allowing detection of the tertiary immune complexes thus formed. This system may provide for signal amplification if this is desired.

The immunodetection methods of the present disclosure have evident utility in the diagnosis of human disease states. A biological or clinical sample suspected of containing either the encoded protein or peptide or corresponding antibody is used. However, these embodiments also have applications to non-clinical samples, such as in the titering of antigen or antibody samples, in the selection of hybridomas, and the like.

In the clinical diagnosis or monitoring of patients with a disease state, the detection of an antigen encoded by a disease state marker nucleic acid, or an increase in the levels of such an antigen, in comparison to the levels in a corresponding biological sample from a normal subject is indicative of a patient with the disease state. The basis for such diagnostic methods lies, in part, with the finding that the nucleic acid disease state markers identified in the present disclosure are overexpressed in peripheral blood samples from individuals with the disease state (see Examples 1 through 4 below). By extension, it may be inferred that at least some of these markers produce elevated levels of encoded proteins, that may also be used as disease state markers.

Methods of differentiating between significant expression of a biomarker, which represents a positive identification, and low level or background expression of a biomarker are well known in the art. Background expression levels are often used to form a "cut-off" above which increased staining is scored as significant or positive. Significant expression may be represented by high levels of antigens in tissues or within body fluids, or alternatively, by a high proportion of cells from within a tissue that each give a positive signal.

2. Immunohistochemistry

The antibodies of the present disclosure may also be used in conjunction with both fresh-frozen and formalin-fixed, paraffin-embedded tissue blocks prepared from study by immunohistochemistry (IHC) or fixed cells on microscope slides for immunocytochemistry. The method of preparing tissue blocks from these particulate specimens has been successfully used in previous IHC studies of various prognostic factors and is well known to those of skill in the art (Brown *et al.*, 1990; Abbondanzo *et al.*, 1990; Allred *et al.*, 1990).

Briefly, frozen-sections may be prepared by rehydrating 50 ng of frozen "pulverized" tissue at room temperature in phosphate buffered saline (PBS) in small plastic capsules; pelleting the particles by centrifugation; resuspending them in a viscous embedding medium (OCT); inverting the capsule and pelleting again by centrifugation; snap-freezing in -70°C isopentane; cutting the plastic capsule and removing the frozen cylinder of tissue; securing the tissue cylinder on a cryostat microtome chuck; and cutting 25-50 serial sections containing an average of about 500 intact cells.

Permanent-sections may be prepared by a similar method involving rehydration of the 50 mg sample in a plastic microfuge tube; pelleting; resuspending in 10% formalin for 4 hours fixation; washing/pelleting; resuspending in warm 2.5% agar; pelleting; cooling in ice water to harden the agar; removing the tissue/agar block from the tube; infiltrating and embedding the block in paraffin; and cutting up to 50 serial permanent sections.

3. Flow Cytometry

Expressed proteins may also be detected by flow cytometry as described in Fujishima et al, 1996. In the practice of the method, the cells are fixed and then incubated with a monoclonal antibody against the expressed protein to be detected. The bound antibodies are then contacted with labeled anti-IgG for example for detection. A typical label is FITC. The fluorescent intensity may then be measured by flow cytometer such as Ortho Cytron, Ortho diagnostics, or FACScan; Becton Dickinson.

FACS permits the separation of sub-populations of cells initially on the basis of their light scatter properties as they pass through a laser beam. The forward light scatter (FALS) is related to cell size and the right angle light scatter to cell density, cell contour and nucleo-cytoplasmic ratio.

Since cells are tagged with fluorescent labeled antibody they can then be further characterized by fluorescence intensity and positive and negative windows set on the FACS to collect bright fluorescence and low fluorescence cells. Cells are sorted at a flow rate of about 3000 cells per second and collected in positive and negative cells.

4. ELISA

As noted, it is contemplated that the encoded proteins or peptides of the disclosure have utility as immunogens, e.g., in connection with vaccine development, in immunohistochemistry and in ELISA assays. One evident utility of the encoded antigens and corresponding antibodies is in immunoassays for the detection of disease state marker proteins, as needed in diagnosis and prognostic monitoring.

Immunoassays, in their most simple and direct sense, are binding assays. Certain preferred immunoassays are the various types of enzyme linked immunosorbent assays (ELISAs) and radioimmunoassays (RIA) known in the art. Immunohistochemical detection using tissue sections is also particularly useful. However, it is readily appreciated that detection is not limited to such techniques, and Western blotting, dot blotting, FACS analyses, and the like may also be used.

In one exemplary ELISA, antibodies binding to the encoded proteins of the disclosure are immobilized onto a selected surface exhibiting protein affinity, such as a well in a polystyrene microtiter plate. Then, a test composition suspected of containing the disease state marker antigen, such as a clinical sample, is added to the wells. After binding and washing to remove non-specifically bound immunecomplexes, the bound antigen may be detected. Detection is generally achieved by the addition of a second antibody specific for the target protein, that is linked to a detectable label. This type of ELISA is a simple "sandwich ELISA". Detection may also be achieved by the addition of a second antibody, followed by the addition of a third antibody that has binding affinity for the second antibody, with the third antibody being linked to a detectable label.

In another exemplary ELISA, the samples suspected of containing the disease state marker antigen are immobilized onto the well surface and then contacted with the antibodies of the invention. After binding and washing to remove non-specifically bound immunecomplexes, the bound antigen is detected. Where the initial antibodies are linked to a detectable label, the immunecomplexes may be detected directly. Again, the immunecomplexes may be detected using

a second antibody that has binding affinity for the first antibody, with the second antibody being linked to a detectable label.

Another ELISA in which the proteins or peptides are immobilized, involves the use of antibody competition in the detection. In this ELISA, labeled antibodies are added to the wells, allowed to bind to the disease state marker protein, and detected by means of their label. The amount of marker antigen in an unknown sample is then determined by mixing the sample with the labeled antibodies before or during incubation with coated wells. The presence of marker antigen in the sample acts to reduce the amount of antibody available for binding to the well and thus reduces the ultimate signal. This is appropriate for detecting antibodies in an unknown sample, where the unlabeled antibodies bind to the antigen-coated wells and reduces the amount of antigen available to bind the labeled antibodies.

Irrespective of the format employed, ELISAs have certain features in common, such as coating, incubating or binding, washing to remove non-specifically bound species, and detecting the bound immunecomplexes. These are described as follows:

In coating a plate with either antigen or antibody, it is typical to incubate the wells of the plate with a solution of the antigen or antibody, either overnight or for a specified period of hours. The wells of the plate are then washed to remove incompletely adsorbed material. Any remaining available surfaces of the wells are then "coated" with a nonspecific protein that is antigenically neutral with regard to the test antisera. These include bovine serum albumin (BSA), casein and solutions of milk powder. The coating allows for blocking of nonspecific adsorption sites on the immobilizing surface and thus reduces the background caused by nonspecific binding of antisera onto the surface.

In ELISAs, it is more customary to use a secondary or tertiary detection means rather than a direct procedure. Thus, after binding of a protein or antibody to the well, coating with a non-reactive material to reduce background, and washing to remove unbound material, the immobilizing surface is contacted with the control and/or clinical or biological sample to be tested under conditions effective to allow immunecomplex (antigen/antibody) formation. Detection of the immunecomplex then requires a labeled secondary binding ligand or antibody, or a secondary binding ligand or antibody in conjunction with a labeled tertiary antibody or third binding ligand.

"Under conditions effective to allow immunecomplex (antigen/antibody) formation" means that the conditions preferably include diluting the antigens and antibodies with solutions such as BSA, bovine gamma globulin (BGG) and phosphate buffered saline (PBS)/Tween. These added agents also tend to assist in the reduction of nonspecific background.

The "suitable" conditions also mean that the incubation is at a temperature and for a period of time sufficient to allow effective binding. Incubation steps are typically from about 1 to 2 to 4 hours, at temperatures preferably on the order of 25° to 27°C, or may be overnight at about 4°C or so.

Following all incubation steps in an ELISA, the contacted surface is washed so as to remove non-complexed material. A preferred washing procedure includes washing with a solution such as PBS/Tween, or borate buffer. Following the formation of specific immunecomplexes between the test sample and the originally bound material, and subsequent washing, the occurrence of even minute amounts of immunecomplexes may be determined.

To provide a detecting means, the second or third antibody has an associated label to allow detection. Preferably, this is an enzyme that generates color development upon incubating with an appropriate chromogenic substrate. Thus, for example, one may contact and incubate the first or second immunecomplex with a urease, glucose oxidase, alkaline phosphatase or hydrogen peroxidase-conjugated antibody for a period of time and under conditions that favor the development of further immunecomplex formation (e.g., incubation for 2 hours at room temperature in a PBS-containing solution such as PBS-Tween).

After incubation with the labeled antibody, and subsequent to washing to remove unbound material, the amount of label is quantified, e.g., by incubation with a chromogenic substrate such as urea and bromocresol purple or 2,2'-azido-di-(3-ethyl)-benzthiazoline-6-sulfonicacid [ABTS] and H_2O_2, in the case of peroxidase as the enzyme label. Quantitation is then achieved by measuring the degree of color generation, *e.g.*, using a spectrophotometer.

5. Use of Antibodies for Radioimaging

The antibodies of this disclosure are used to quantify and localize the expression of the encoded marker proteins. The antibody, for example, may be labeled by any one of a variety of

methods and used to visualize the localized concentration of the cells producing the encoded protein.

A radionuclide may be bound to an antibody either directly or indirectly by using an intermediary functional group. Intermediary functional groups which are often used to bind radioisotopes which exist as metallic ions to antibody are diethylenetriaminepentaacetic acid (DTPA) and ethylene diaminetetracetic acid (EDTA). Examples of metallic ions suitable for use in this disclosure are ^{99m}Tc, ^{123}I, ^{131}I ^{111}In, ^{97}Ru, ^{67}Cu, ^{67}Ga, ^{125}I, ^{68}Ga, ^{72}As, ^{89}Zr, and ^{201}Tl.

In accordance with this disclosure, the monoclonal antibody or fragment thereof may be labeled by any of several techniques known to the art. The methods of the present disclosure may also use paramagnetic isotopes for purposes of *in vivo* detection. Elements particularly useful in Magnetic Resonance Imaging ("MRI") include ^{157}Gd, ^{55}Mn, ^{162}Dy, ^{52}Cr, and ^{56}Fe.

Administration of the labeled antibody may be local or systemic and accomplished intravenously, intraarterially, via the spinal fluid or the like. Administration may also be intradermal or intracavitary, depending upon the body site under examination. After a sufficient time has lapsed for the monoclonal antibody or fragment thereof to bind with the diseased tissue, for example 30 minutes to 48 hours, the area of the subject under investigation is examined by routine imaging techniques such as MRI, SPECT, planar scintillation imaging and emerging imaging techniques, as well. The exact protocol necessarily varies depending upon factors specific to the patient, as noted above, and depending upon the body site under examination, method of administration and type of label used. The determination of specific procedures is routine in the art. The distribution of the bound radioactive isotope and its increase or decrease with time is then monitored and recorded. By comparing the results with data obtained from studies of clinically normal individuals, the presence and extent of the diseased tissue may be determined.

The instant disclosure addresses detection of disease state cells by their effect on gene expression in immune system lymphocytes. In early stages of the disease state, such immune response may be localized. For example, the response may be limited to lymph nodes immediately surrounding a metastasizing tumor or other localized form of a disease state. Localization of differentially expressed disease state markers may be of utility for separating disease states of widespread distribution from those of limited distribution within the patient. Such a detection

means is therefore of significance in the management and care of patients with the disease state. It will be recognized that this utility is included within the scope of the present disclosure.

6. Kits

In still further embodiments, the present disclosure concerns immunodetection kits for use with the immunodetection methods described above. As the encoded proteins or peptides may be employed to detect antibodies and the corresponding antibodies may be employed to detect encoded proteins or peptides, either or both of such components may be provided in the kit. The immunodetection kits thus comprise, in suitable container means, an encoded protein or peptide, or a first antibody that binds to an encoded protein or peptide, and an immunodetection reagent.

In certain embodiments, the encoded protein or peptide, or the first antibody that binds to the encoded protein or peptide, may be bound to a solid support, such as a column matrix or well of a microtiter plate.

The immunodetection reagents of the kit may take any one of a variety of forms, including those detectable labels that are associated with or linked to the given antibody or antigen, and detectable labels that are associated with or attached to a secondary binding ligand. Exemplary secondary ligands are those secondary antibodies that have binding affinity for the first antibody or antigen, and secondary antibodies that have binding affinity for a human antibody.

Further suitable immunodetection reagents for use in the present kits include the two-component reagent that comprises a secondary antibody that has binding affinity for the first antibody or antigen, along with a third antibody that has binding affinity for the second antibody, the third antibody being linked to a detectable label.

The kits may further comprise a suitably aliquoted composition of the encoded protein or polypeptide antigen, whether labeled or unlabeled, as may be used to prepare a standard curve for a detection assay.

The kits may contain antibody-label conjugates either in fully conjugated form, in the form of intermediates, or as separate moieties to be conjugated by the user of the kit. The components of the kits may be packaged either in aqueous media or in lyophilized form.

The container means of the kits generally includes at least one vial, test tube, flask, bottle, syringe or other container means, into which the antibody or antigen may be placed, and preferably,

suitably aliquoted. Where a second or third binding ligand or additional component is provided, the kit also generally contains a second, third or other additional container into which this ligand or component may be placed. The kits of the present disclosure also typically include a means for containing the antibody, antigen, and any other reagent containers in close confinement for commercial sale. Such containers may include injection or blow-molded plastic containers into which the desired vials are retained.

E. Detection and Quantitation of RNA Species

One embodiment of the instant disclosure comprises a method for identification of a disease state in a biological sample by amplifying and detecting nucleic acids corresponding to disease state markers. The biological sample may be any tissue or fluid in which lymphocyte cells might be present. Various embodiments include bone marrow aspirate, bone marrow biopsy, lymph node aspirate, lymph node biopsy, spleen tissue, fine needle aspirate, skin biopsy or organ tissue biopsy. Other embodiments include samples of body fluid such as peripheral blood, lymph fluid, ascites, serous fluid, pleural effusion, sputum, cerebrospinal fluid, lacrimal fluid, stool or urine.

Nucleic acid used as a template for amplification is isolated from cells contained in the biological sample, according to conventional methodologies. (Sambrook *et al.*, 1989) The nucleic acid may be genomic DNA or fractionated or whole cell RNA. Where RNA is used, it may be desired to convert the RNA to a complementary cDNA. In one embodiment, the RNA is whole cell RNA and is used directly as the template for amplification.

Pairs of primers that selectively hybridize to nucleic acids corresponding to disease state-specific markers are contacted with the isolated nucleic acid under conditions that permit selective hybridization. Once hybridized, the nucleic acid:primer complex is contacted with one or more enzymes that facilitate template-dependent nucleic acid synthesis. Multiple rounds of amplification, also referred to as "cycles," are conducted until a sufficient amount of amplification product is produced.

Next, the amplification product is detected. In certain applications, the detection may be performed by visual means. Alternatively, the detection may involve indirect identification of the product via chemiluminescence, radioactive scintigraphy of incorporated radiolabel or fluorescent

label or even via a system using electrical or thermal impulse signals (Affymax technology; Bellus, 1994).

Following detection, one may compare the results seen in a given patient with statistically significant reference groups of normal individuals and patients with the disease state. In this way, it is possible to correlate the amount of marker detected with various clinical states.

1. Primers

The term primer, as defined herein, is meant to encompass any nucleic acid that is capable of priming the synthesis of a nascent nucleic acid in a template-dependent process. Typically, primers are oligonucleotides from ten to twenty base pairs in length, but longer sequences may be employed. Primers may be provided in double-stranded or single-stranded form, although the single-stranded form is preferred.

2. Template Dependent Amplification Methods

A number of template dependent processes are available to amplify the marker sequences present in a given template sample. One of the best known amplification methods is the polymerase chain reaction (referred to as PCR) which is described in detail in U.S. Patent Nos. 4,683,195, 4,683,202 and 4,800,159, and in Innis *et al.*, 1990, each of which is incorporated herein by reference in its entirety.

Briefly, in PCR, two primer sequences are prepared which are complementary to regions on opposite complementary strands of the marker sequence. An excess of deoxynucleoside triphosphates is added to a reaction mixture along with a DNA polymerase, *e.g.*, *Taq* polymerase. If the marker sequence is present in a sample, the primers bind to the marker and the polymerase causes the primers to be extended along the marker sequence by adding on nucleotides. By raising and lowering the temperature of the reaction mixture, the extended primers dissociate from the marker to form reaction products, excess primers bind to the marker and to the reaction products and the process is repeated.

A reverse transcriptase PCR amplification procedure may be performed in order to quantify the amount of mRNA amplified. Methods of reverse transcribing RNA into cDNA are well known and described in Sambrook *et al.*, 1989. Alternative methods for reverse transcription utilize

thermostable DNA polymerases. These methods are described in WO 90/07641 filed December 21, 1990. Polymerase chain reaction methodologies are well known in the art.

Alternatively, RNA species can be quantitated by means that do not necessarily require amplification by PCR. These means may include other amplification techniques, for example, isothermic amplification techniques such as the one developed by Gen-Probe (San Diego, CA), and the ligase chain reaction ("LCR"), disclosed in EPA No. 320 308, incorporated herein by reference in its entirely. In LCR, two complementary probe pairs are prepared, and in the presence of the target sequence, each pair binds to opposite complementary strands of the target such that they abut. In the presence of a ligase, the two probe pairs link to form a single unit. By temperature cycling, as in PCR, bound ligated units dissociate from the target and then serve as "target sequences" for ligation of excess probe pairs. U.S. Patent 4,883,750 describes a method similar to LCR for binding probe pairs to a target sequence.

Qbeta Replicase, described in PCT Application No. PCT/US87/00880, may also be used as still another amplification method in the present invention. In this method, a replicative sequence of RNA which has a region complementary to that of a target is added to a sample in the presence of an RNA polymerase. The polymerase copies the replicative sequence which may then be detected.

An isothermal amplification method, in which restriction endonucleases and ligases are used to achieve the amplification of target molecules that contain nucleoside 5'-[alpha-thio]-triphosphates in one strand of a restriction site may also be useful in the amplification of nucleic acids in the present invention. Walker *et al.*, *Proc. Nat'l Acad. Sci. USA* 89:392-396 (1992), incorporated herein by reference in its entirety.

Strand Displacement Amplification (SDA) is another method of carrying out isothermal amplification of nucleic acids which involves multiple rounds of strand displacement and synthesis, *i.e.*, nick translation. A similar method, called Repair Chain Reaction (RCR), involves annealing several probes throughout a region targeted for amplification, followed by a repair reaction in which only two of the four bases are present. The other two bases may be added as biotinylated derivatives for easy detection. A similar approach is used in SDA. Target specific sequences may also be detected using a cyclic probe reaction (CPR). In CPR, a probe having 3' and 5' sequences of non-specific DNA and a middle sequence of specific RNA is hybridized to DNA which is present

in a sample. Upon hybridization, the reaction is treated with RNase H, and the products of the probe identified as distinctive products which are released after digestion. The original template is annealed to another cycling probe and the reaction is repeated.

Other amplification methods are described in GB Application No. 2 202 328, and in PCT Application No. PCT/US89/01025, each of which is incorporated herein by reference in its entirety, may be used in accordance with the present invention. In the former application, "modified" primers are used in a PCR like, template and enzyme dependent synthesis. The primers may be modified by labeling with a capture moiety (*e.g.*, biotin) and/or a detector moiety (*e.g.*, enzyme). In the latter application, an excess of labeled probes are added to a sample. In the presence of the target sequence, the probe binds and is cleaved catalytically. After cleavage, the target sequence is released intact to be bound by excess probe. Cleavage of the labeled probe signals the presence of the target sequence.

Other nucleic acid amplification procedures include transcription-based amplification systems (TAS), including nucleic acid sequence based amplification (NASBA) and 3SR. Kwoh *et al.*, *Proc. Nat'l Acad. Sci. USA* 86:1173 (1989); Gingeras *et al.*, PCT Application WO 88/10315, incorporated herein by reference in their entirety. In NASBA, the nucleic acids may be prepared for amplification by conventional phenol/chloroform extraction, heat denaturation of a clinical sample, treatment with lysis buffer and minispin columns for isolation of DNA and RNA or guanidinium chloride extraction of RNA. These amplification techniques involve annealing a primer which has target specific sequences. Following polymerization, DNA/RNA hybrids are digested with RNase H while double stranded DNA molecules are heat denatured again. In either case the single stranded DNA is made fully double stranded by addition of second target specific primer, followed by polymerization. The double-stranded DNA molecules are then multiply transcribed by a polymerase such as T7 or SP6. In an isothermal cyclic reaction, the RNA's are reverse transcribed into double stranded DNA, and transcribed once against with a polymerase such as T7 or SP6. The resulting products, whether truncated or complete, indicate target specific sequences.

Davey *et al.*, EPA No. 329 822 (incorporated herein by reference in its entirely) disclose a nucleic acid amplification process involving cyclically synthesizing single-stranded RNA ("ssRNA"), ssDNA, and double-stranded DNA (dsDNA), which may be used in accordance with

the present invention. The ssRNA is a first template for a first primer oligonucleotide, which is elongated by reverse transcriptase (RNA-dependent DNA polymerase). The RNA is then removed from the resulting DNA:RNA duplex by the action of ribonuclease H (RNase H, an RNase specific for RNA in duplex with either DNA or RNA). The resultant ssDNA is a second template for a second primer, which also includes the sequences of an RNA polymerase promoter (exemplified by T7 RNA polymerase) 5' to its homology to the template. This primer is then extended by DNA polymerase (exemplified by the large "Klenow" fragment of *E. coli* DNA polymerase I), resulting in a double-stranded DNA ("dsDNA") molecule, having a sequence identical to that of the original RNA between the primers and having additionally, at one end, a promoter sequence. This promoter sequence may be used by the appropriate RNA polymerase to make many RNA copies of the DNA. These copies may then re-enter the cycle leading to very swift amplification. With proper choice of enzymes, this amplification may be done isothermally without addition of enzymes at each cycle. Because of the cyclical nature of this process, the starting sequence may be chosen to be in the form of either DNA or RNA.

Miller *et al.*, PCT Application WO 89/06700 (incorporated herein by reference in its entirety) disclose a nucleic acid sequence amplification scheme based on the hybridization of a promoter/primer sequence to a target single-stranded DNA ("ssDNA") followed by transcription of many RNA copies of the sequence. This scheme is not cyclic, *i.e.*, new templates are not produced from the resultant RNA transcripts. Other amplification methods include "race" and "one-sided PCR." Frohman, M.A., In: *PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS*, Academic Press, N.Y. (1990) and Ohara *et al.*, *Proc. Nat'l Acad. Sci. USA*, 86:5673-5677 (1989), each herein incorporated by reference in their entirety.

Methods based on ligation of two (or more) oligonucleotides in the presence of nucleic acid having the sequence of the resulting "di-oligonucleotide", thereby amplifying the di-oligonucleotide, may also be used in the amplification step of the present invention. Wu *et al.*, *Genomics* 4:560 (1989), incorporated herein by reference in its entirety.

An example of a technique that does not require nucleic acid amplification, that can also be used to quantify RNA in some applications is a nuclease protection assay. There are many different versions of nuclease protection assays known to those practiced in the art. The characteristic that all versions of nuclease protection assays share in common is that they involve hybridization of an

antisense nucleic acid with the RNA to be quantified. The resulting hybrid double stranded molecule is then digested with a nuclease that digests single stranded nucleic acids more efficiently than double stranded molecules. The amount of antisense nucleic acid that survives digestion is a measure of the amount of the target RNA species to be quantified. An example of a nuclease protection assay that is commercially available is the RNase protection assay manufactured by Ambion, Inc. (Austin, TX).

3. Separation Methods

Following amplification, it may be desirable to separate the amplification product from the template and the excess primer for the purpose of determining whether specific amplification has occurred. In one embodiment, amplification products are separated by agarose, agarose-acrylamide or polyacrylamide gel electrophoresis using conventional methods. See Sambrook *et al.*, 1989.

Alternatively, chromatographic techniques may be employed to effect separation. There are many kinds of chromatography which may be used in the present invention: adsorption, partition, ion-exchange and molecular sieve, HPLC, and many specialized techniques for using them including column, paper, thin-layer and gas chromatography (Freifelder, 1982).

Another example of a separation methodology is done by covalently labeling the oligonucleotide primers used in a PCR reaction with various types of small molecule ligands. In one such separation, a different ligand is present on each oligonucleotide. A molecule, perhaps an antibody or avidin if the ligand is biotin, that specifically binds to one of the ligands is used to coat the surface of a plate such as a 96 well ELISA plate. Upon application of the PCR reactions to the surface of such a prepared plate, the PCR products are bound with specificity to the surface. After washing the plate to remove unbound reagents, a solution containing a second molecule that binds to the first ligand is added. This second molecule is linked to some kind of reporter system. The second molecule only binds to the plate if a PCR product has been produced whereby both oligonucleotide primers are incorporated into the final PCR products. The amount of the PCR product is then detected and quantified in a commercial plate reader much as ELISA reactions are detected and quantified. An ELISA-like system such as the one

described here has been developed by the Raggio Italgene company under the C-Track trade name.

4. Identification Methods

Amplification products must be visualized in order to confirm amplification of the marker sequences. One typical visualization method involves staining of a gel with ethidium bromide and visualization under UV light. Alternatively, if the amplification products are integrally labeled with radio- or fluorometrically-labeled nucleotides, the amplification products may then be exposed to x-ray film or visualized under the appropriate stimulating spectra, following separation.

In one embodiment, visualization is achieved indirectly. Following separation of amplification products, a labeled, nucleic acid probe is brought into contact with the amplified marker sequence. The probe preferably is conjugated to a chromophore but may be radiolabeled. In another embodiment, the probe is conjugated to a binding partner, such as an antibody or biotin, where the other member of the binding pair carries a detectable moiety.

In one embodiment, detection is by Southern blotting and hybridization with a labeled probe. The techniques involved in Southern blotting are well known to those of skill in the art and may be found in many standard books on molecular protocols. See Sambrook *et al.*, 1989. Briefly, amplification products are separated by gel electrophoresis. The gel is then contacted with a membrane, such as nitrocellulose, permitting transfer of the nucleic acid and non-covalent binding. Subsequently, the membrane is incubated with a chromophore-conjugated probe that is capable of hybridizing with a target amplification product. Detection is by exposure of the membrane to x-ray film or ion-emitting detection devices.

One example of the foregoing is described in U.S. Patent No. 5,279,721, incorporated by reference herein, which discloses an apparatus and method for the automated electrophoresis and transfer of nucleic acids. The apparatus permits electrophoresis and blotting without external manipulation of the gel and is ideally suited to carrying out methods according to the present invention.

5. Kit Components

All the essential materials and reagents required for detecting disease state markers in a biological sample may be assembled together in a kit. This generally comprises preselected primers for specific markers. Also included may be enzymes suitable for amplifying nucleic acids including various polymerases (RT, Taq, etc.), deoxynucleotides and buffers to provide the necessary reaction mixture for amplification.

Such kits generally comprise, in suitable means, distinct containers for each individual reagent and enzyme as well as for each marker primer pair. Preferred pairs of primers for amplifying nucleic acids are selected to amplify the sequences specified in Genebank Accession numbers D87451, T03013, X03558, M28130, Y00787, SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:29.

In another embodiment, such kits comprise hybridization probes specific for disease state markers, chosen from a group including nucleic acids corresponding to the sequences specified in Genebank Accession numbers D87451, T03013, X03558, M28130, Y00787, SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:29. Such kits generally comprise, in suitable means, distinct containers for each individual reagent and enzyme as well as for each marker hybridization probe.

F. Use of RNA Fingerprinting to Identify Markers of Human Disease

RNA fingerprinting is a means by which RNAs isolated from many different tissues, cell types or treatment groups may be sampled simultaneously to identify RNAs whose relative abundances vary. Two forms of this technology were developed simultaneously and reported in 1992 as RNA fingerprinting by differential display (Liang and Pardee, 1992; Welsh *et al.*, 1992). (*See also* Liang and Pardee, U.S. Patent 5,262,311, incorporated herein by reference in its entirety.) Some of the studies described herein were performed similarly to Donahue *et al.*, *J. Biol. Chem.* 269: 8604-8609, 1994.

All forms of RNA fingerprinting by PCR are theoretically similar but differ in their primer design and application. The most striking difference between differential display and other methods of RNA fingerprinting is that differential display utilizes anchoring primers that hybridize

to the poly A tails of mRNAs. As a consequence, the PCR products amplified in differential display are biased towards the 3' untranslated regions of mRNAs.

The basic technique of differential display has been described in detail (Liang and Pardee, 1992). Total cell RNA is primed for first strand reverse transcription with an anchoring primer composed of oligo dT. The oligo dT primer is extended using a reverse transcriptase, for example, Moloney Murine Leukemia Virus (MMLV) reverse transcriptase. The synthesis of the second strand is primed with an arbitrarily chosen oligonucleotide, using reduced stringency conditions. Once the double-stranded cDNA has been synthesized, amplification proceeds by conventional PCR techniques, utilizing the same primers. The resulting DNA fingerprint is analyzed by gel electrophoresis and ethidium bromide staining or autoradiography. A side by side comparison of fingerprints obtained from different cell derived RNAs using the same oligonucleotide primers identifies mRNAs that are differentially expressed.

RNA fingerprinting technology has been demonstrated as being effective in identifying genes that are differentially expressed in cancer cells (Liang *et al.*, 1992; Wong *et al.*, 1993; Sager *et al.*, 1993; Mok *et al.*, 1994; Watson *et al.*, 1994; Chen *et al.*, 1995; An *et al.*, 1995). The present disclosure utilizes the RNA fingerprinting technique or other techniques described herein to identify genes that are differentially expressed in peripheral blood cells in human disease states.

G. Design and Theoretical Considerations for Relative Quantitative RT-PCR

Reverse transcription (RT) of RNA to cDNA followed by relative quantitative PCR (RT-PCR) may be used to determine the relative concentrations of specific mRNA species in a series of total cell RNAs isolated from peripheral blood of normal individuals and individuals with a disease state. By determining that the concentration of a specific mRNA species varies, it is shown that the gene encoding the specific mRNA species is differentially expressed. This technique may be used to confirm that mRNA transcripts shown to be differentially regulated by RNA fingerprinting are differentially expressed in disease state progression.

In PCR, the number of molecules of the amplified target DNA increase by a factor approaching two with every cycle of the reaction until some reagent becomes limiting. Thereafter, the rate of amplification becomes increasingly diminished until there is not an increase in the amplified target between cycles. If one plots a graph on which the cycle number is on the X axis

and the log of the concentration of the amplified target DNA is on the Y axis, one observes that a curved line of characteristic shape is formed by connecting the plotted points. Beginning with the first cycle, the slope of the line is positive and constant. This is said to be the linear portion of the curve. After some reagent becomes limiting, the slope of the line begins to decrease and eventually becomes zero. At this point the concentration of the amplified target DNA becomes asymptotic to some fixed value. This is said to be the plateau portion of the curve.

The concentration of the target DNA in the linear portion of the PCR is directly proportional to the starting concentration of the target before the PCR was begun. By determining the concentration of the PCR products of the target DNA in PCR reactions that have completed the same number of cycles and are in their linear ranges, it is possible to determine the relative concentrations of the specific target sequence in the original DNA mixture. If the DNA mixtures are cDNAs synthesized from RNAs isolated from different tissues or cells, the relative abundances of the specific mRNA from which the target sequence was derived may be determined for the respective tissues or cells. This direct proportionality between the concentration of the PCR products and the relative mRNA abundances is only true in the linear range portion of the PCR reaction.

The final concentration of the target DNA in the plateau portion of the curve is determined by the availability of reagents in the reaction mix and is independent of the original concentration of target DNA. Therefore, the first condition that must be met before the relative abundances of a mRNA species may be determined by RT-PCR for a collection of RNA populations is that the concentrations of the amplified PCR products must be sampled when the PCR reactions are in the linear portion of their curves.

The second condition that must be met for an RT-PCR study to successfully determine the relative abundances of a particular mRNA species is that relative concentrations of the amplifiable cDNAs must be normalized to some independent standard. The goal of an RT-PCR study is to determine the abundance of a particular mRNA species relative to the average abundance of all mRNA species in the sample. In the studies described below, mRNAs for ß-actin, asparagine synthetase and lipocortin II were used as external and internal standards to which the relative abundance of other mRNAs are compared.

Most protocols for competitive PCR utilize internal PCR standards that are approximately as abundant as the target. These strategies are effective if the products of the PCR amplifications are sampled during their linear phases. If the products are sampled when the reactions are approaching the plateau phase, then the less abundant product becomes relatively over represented. Comparisons of relative abundances made for many different RNA samples, such as is the case when examining RNA samples for differential expression, become distorted in such a way as to make differences in relative abundances of RNAs appear less than they actually are. This is not a significant problem if the internal standard is much more abundant than the target. If the internal standard is more abundant than the target, then direct linear comparisons may be made between RNA samples.

The discussion above describes the theoretical considerations for an RT-PCR assay for clinically derived materials. The problems inherent in clinical samples are that they are of variable quantity (making normalization problematic), and that they are of variable quality (necessitating the co-amplification of a reliable internal control, preferably of larger size than the target). Both of these problems are overcome if the RT-PCR is performed as a relative quantitative RT-PCR with an internal standard in which the internal standard is an amplifiable cDNA fragment that is larger than the target cDNA fragment and in which the abundance of the mRNA encoding the internal standard is roughly 5-100 fold higher than the mRNA encoding the target. This assay measures relative abundance, not absolute abundance of the respective mRNA species.

Other studies may be performed using a more conventional relative quantitative RT-PCR with an external standard protocol. These assays sample the PCR products in the linear portion of their amplification curves. The number of PCR cycles that are optimal for sampling must be empirically determined for each target cDNA fragment. In addition, the reverse transcriptase products of each RNA population isolated from the various tissue samples must be carefully normalized for equal concentrations of amplifiable cDNAs. While empirical determination of the linear range of the amplification curve and normalization of cDNA preparations are tedious and time consuming processes, the resulting RT-PCR assays may, in certain cases, be superior to those derived from a relative quantitative RT-PCR with an internal standard.

One reason for this is that without the internal standard/competitor, all of the reagents may be converted into a single PCR product in the linear range of the amplification curve, increasing the

sensitivity of the assay. Another reason is that with only one PCR product, display of the product on an electrophoretic gel or some other display method becomes less complex, has less background and is easier to interpret.

H. Diagnosis and Prognosis of Human Cancer

In certain embodiments, the present disclosure enables the diagnosis and prognosis of human cancer by screening for marker nucleic acids. Various markers have been proposed to be correlated with metastasis and malignancy. They may be classified generally as cytologic, protein or nucleic acid markers.

Cytologic markers include such things as "nuclear roundedness" (Diamond *et al.*, 1982) and cell ploidy. Protein markers include prostate specific antigen (PSA) and CA125. Nucleic acid markers have included amplification of Her2/*neu*, point mutations in the p53 or *ras* genes, and changes in the sizes of triplet repeat segments of particular chromosomes.

All of these markers exhibit certain drawbacks, associated with false positives and false negatives. A false positive result occurs when an individual without malignant cancer exhibits the presence of a "cancer marker". For example, elevated serum PSA has been associated with prostate carcinoma. However, it also occurs in some individuals with non-malignant, benign hyperplasia of the prostate. A false negative result occurs when an individual actually has cancer, but the test fails to show the presence of a specific marker. The incidence of false negatives varies for each marker, and frequently also by tissue type. For example, *ras* point mutations have been reported to range from a high of 95 percent in pancreatic cancer to a low of zero percent in some gynecologic cancers.

Additional problems arise when a marker is present only within the transformed cell itself. Ras point mutations may only be detected within the mutant cell, and are apparently not present in, for example, the serum or urine of individuals with *ras*-activated carcinomas. This means that, in order to detect a malignant tumor, one must take a sample of the tumor itself, or its metastatic cells. Essentially one must first identify and sample a tumor before the presence of the cancer marker may be detected.

Finally, specific problems occur with markers that are present in normal cells but absent in cancer cells. Most tumor samples contain mixed populations of both normal and transformed cells.

If one is searching for a marker that is present in normal cells, but occurs at reduced levels in transformed cells, the "background" signal from the normal cells in the sample may mask the presence of transformed cells.

The ideal disease state marker would be one that is present in individuals with the disease state, and either missing or expressed at significantly lower levels in normal individuals. The present disclosure addresses this need, in the case of metastatic prostate cancer for example, by identifying several new nucleic acid markers that are expressed at higher levels in individuals with metastatic prostate cancer than in normal individuals. In particular, the results for markers UC302 (SEQ ID #3) and UC325 (SEQ ID #4) are quite promising in that these markers are apparently only overexpressed in the peripheral blood of individuals with metastatic tumors and are present at relatively low levels in normal individuals.

Further, since the markers are present in the whole blood of individuals with the disease state, the present detection method avoids the problem of having to suspect a tumor is in place before it may be sampled. The instant disclosure has utility as a general screening tool for asymptomatic individuals, as well as a means of differentially diagnosing those patients whose tumors have already metastasized. Depending upon the type of tumor involved, such individuals may be selected for systemic forms of anti-cancer therapy rather than surgical removal of localized tumor masses. Certain individuals with advanced forms of highly malignant metastatic tumors may be optimally treated by pain management alone.

It is anticipated that in clinical applications, human tissue samples will be screened for the presence of the disease state markers identified herein. Such samples would normally consist of peripheral blood, but may also consist of needle biopsy cores or lymph node tissue. In certain embodiments, nucleic acids would be extracted from these samples and amplified as described above. Some embodiments would utilize kits containing pre-selected primer pairs or hybridization probes. The amplified nucleic acids would be tested for the markers by, for example, gel electrophoresis and ethidium bromide staining, or Southern blotting, or a solid-phase detection means as described above. These methods are well known within the art. The levels of selected markers detected would be compared with statistically valid groups of individuals with metastatic, non-metastatic malignant, or benign tumors or normal individuals. The diagnosis and prognosis of the individual patient would be determined by comparison with such groups.

Another embodiment of the present disclosure involves application of RT-PCR techniques to detect a disease state using probes and primers selected from sequences comprising Genebank Accession numbers D87451, T03013, X03558, M28130, Y00787, SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:29. Similar techniques have been described in PCT Patent Application No. WO 94/10343, incorporated herein by reference.

In this embodiment, the disease state is detected in hematopoietic samples by amplification of disease state-specific nucleic acid sequences. Samples taken from blood or lymph nodes are treated as described below to purify total cell RNA. The isolated RNA is reverse transcribed using a reverse transcriptase and primers selected to bind under high stringency conditions to a nucleic acid sequence from a group comprising Genebank Accession numbers D87451, T03013, X03558, M28130, Y00787, SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:29. Following reverse transcription, the resulting cDNAs are amplified using conventional PCR techniques and a thermostable DNA polymerase.

The presence of amplification products corresponding to disease state-marker nucleic acids may be detected by several alternative means. In one embodiment, the amplification product may be detected by gel electrophoresis and ethidium bromide staining. Alternatively, following the gel electrophoresis step the amplification product may be detected by conventional Southern blotting techniques, using an hybridization probe selected to bind specifically to a disease state-marker nucleic acid sequence. Probe hybridization may in turn be detected by a conventional labeling means, for example, by incorporation of [^{32}P]-nucleotides followed by autoradiography. The amplification products may alternatively be detected using a solid phase detection system such as those utilizing a disease state-marker specific hybridization probe and an appropriate labeling means, or even the ELISA-like system known as C-track™ as described above.

The following examples are included to demonstrate preferred embodiments of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventors to function well in the practice of the invention, and thus may be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes may be made in the particular embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

I. Materials and Methods

1. Application of RNA fingerprinting to discover biomarkers for disease states

RNA fingerprinting (according to Liang and Pardee, 1992; Welsh *et al.*, 1992; Liang and Pardee, 1993) was applied to nucleic acids isolated from the peripheral blood of individuals with metastatic prostate cancer, compared with normal individuals.

Blood was drawn from cancer patients and normal individuals into Vacutainer CPT tubes with ficol gradients (Becton Dickinson and Company, Frankin Lanes, NJ). The tubes were centrifuged to separate the red blood cells from various types of nucleated cells, collectively referred to as the buffy coat, and from blood plasma. Total cell RNA was isolated from the buffy coats by the RNA STAT-60 method (Tel-Test, Inc., Friendswood, TX). After RNA isolation, the nucleic acids were precipitated with ethanol. The precipitates were pelleted by centrifugation and redissolved in water. The redissolved nucleic acids were then digested with RNase-free DNase I (Boehringer Mannheim, Inc.) following the manufacturer's instructions, followed by organic extraction with phenol:chloroform:isoamylalcohol (25:24:1) and re-precipitation with ethanol.

The DNase I treated RNA was then pelleted by centrifugation and redissolved in water. The purity and concentration of the RNA in solution was estimated by determining optical density at wave lengths of 260 nm and 280 nm (Sambrook *et al.*, 1989). The RNA was then examined by electrophoresis on a native TAE agarose gel (Sambrook *et al.*, 1989) to determine its integrity. The RNA was then divided into three aliquots. One aliquot was set aside for relative quantitative RT-PCR confirmation using the external standard method described below.

A second aliquot was used to fingerprint the RNA by converting the RNA to first strand cDNA using random hexamers and reverse transcriptase; fingerprinting the cDNA by PCR using arbitrarily chosen oligonucleotides, (10 nucleotides in length); displaying the resulting PCR amplified products on an agarose gel stained with ethidium bromide and cutting differentially appearing bands out of the gel. The excised bands were then cloned and sequenced.

The RNA of the third aliquot was pooled to make a pool of blood RNA from normal individuals and a pool of RNA from the blood of patients with metastatic prostate cancer. The pools were fingerprinted using the sequential pairwise method of arbitrarily primed PCR fingerprinting of RNA (McClelland *et al.*, 1994, *Nucleic Acids Research* 22, 4419-4431,

incorporated herein by reference) with several changes. For example, arbitrary oligonucleotides of 15 to 24 nucleotides were used with Taq polymerase, and one tenth of each first strand cDNA reaction in each arbitrarily primed PCR reaction. One hundred and 200 ng were used in each first strand cDNA synthesis, respectively. Certain genes disclosed herein were discovered by the sequential pairwise method.

2. Methods Utilized in the RNA Fingerprinting Technique

The second type of RNA fingerprinting studies performed more closely resembled the protocol of Welsh *et al.* (1992). This approach used a variation of the above as modified by the use of agarose gels and non-isotopic detection of bands by ethidium bromide staining (An *et al.*, 1995). Total RNAs were isolated from peripheral blood samples as described (Chomczynski & Sacchi, 1987). Ten micrograms of total cellular RNAs were treated with 5 units of RNAse-free DNAse I (GIBCO/BRL) in 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 2 mM $MgCl_2$, and 20 units of RNAse inhibitor (Boehringer Mannheim). After extraction with phenol/chloroform and ethanol precipitation, the RNAs were redissolved in DEPC-treated water.

Two μg of each total cell RNA sample was reverse transcribed into cDNA using randomly selected hexamer primers and MMLV reverse transcriptase (GIBCO/BRL). PCR was performed using one or two arbitrarily chosen oligonucleotide primers (10-12mers). PCR conditions were: 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 50 mM dNTPs, 0.2 mM of primer(s), 1 unit of Taq DNA polymerase (GIBCO/BRL) in a final volume of 20 ml. The amplification parameters included 35 cycles of reaction with 30 sec denaturing at 94°C, 90 sec annealing at 40°C, and 60 sec extension at 72°C. A final extension at 72°C was performed for 15 min. The resulting PCR products were resolved into a fingerprint by size separation by electrophoresis through 2% agarose gels in TBE buffer (Sambrook *et al.*, 1989). The fingerprints were visualized by staining with ethidium bromide. No re-amplification was performed.

Differentially appearing PCR products, that might represent differentially expressed genes, were excised from the gel with a razor blade, purified from the agarose using the Geneclean kit (Bio 101, Inc.), eluted in water and cloned directly into plasmid vectors using the TA cloning strategy (Invitrogen, Inc., and Promega, Inc.). These products were not re-amplified after the initial PCR fingerprinting protocol.

3. Confirmation of Differential Expression by Relative Quantitative RT-PCR: Protocols for RT-PCR

a. Reverse transcription

One to five μg of total cell RNA from each tissue sample was reverse transcribed into cDNA. Reverse transcription was performed with 400 units of MMLV reverse transcriptase (GIBCO/BRL) in the presence of 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, 500 mM dNTP, 50 ng random hexamers per microgram of RNA, and 1 U/ml RNase inhibitor. The reaction volume was 60 ml. The reaction mixture was incubated at room temperature for 10 minutes, then at 37°C for 50 minutes. After reverse transcription the enzyme was denatured by heating to 65°C for 10 minutes. After heat denaturation the samples were diluted with water to a final volume of 300 ml.

RT-PCR was utilized to examine mRNAs for differential expression. The sequences of oligonucleotides used as primers to direct the amplification of the various cDNA fragments are presented in Table 3.

b. Relative Quantitative RT-PCR With an Internal Standard

The concentrations of the original total cell RNAs were determined by measurement of $OD_{260/280}$ (Sambrook *et al.*, 1989) and confirmed by examination of ribosomal RNAs on ethidium bromide stained agarose gels. It is required that all quantitative PCR reactions be normalized for equal amounts of amplifiable cDNA after the reverse transcription is completed. One solution to this is to terminate the reactions by driving the PCR reactions into plateau phase. This approach was utilized in some studies because it is quick and efficient. Lipocortin II was used as the internal standard or competitor. These PCRs were set up as follows:

Reagents: 200 mM each dNTP, 200 nM each oligonucleotide primer, 1X PCR buffer (Boehringer Mannheim including 1.5 mM $MgCl_2$), 3 ml diluted cDNA, and 2.5 units of Taq DNA polymerase/100ml of reaction volume.

Cycling parameters: 30 cycles of 94°C for 1 min; 55°C for 1 min; and 72°C for two min. Thermocyclers were either the MJ research thermocycler or the Stratagene Robocycler.

c. Relative Quantitative RT-PCR with an External Standard

There are three problems with the relative quantitative RT-PCR strategy described above. First, the internal standard must be roughly 4-10 times more abundant than the target for this strategy to normalize the samples. Second, because most of the PCR products are templated from the more abundant internal standard, the assay is less than optimally sensitive. Third, the internal standard must be truly unvarying. The result is that while the strategy described above is fast, convenient and applicable to samples of varying quality, it lacks sensitivity to modest changes in abundances.

To address these issues, a normalization was performed using the ß-actin mRNA as external standard. These PCR reactions were performed with sufficient cycles to observe the products in the linear range of their amplification curves. The intensities of the ethidium bromide stained bands were documented and quantified using the Is1000 imaging analysis system manufactured by the Alpha Innotech, Corp. The quantified data was then normalized for variations in the starting concentrations of amplifiable cDNA by comparing the quantified data from each study with that derived from a similar study which amplified a cDNA fragment copied from the ß-actin mRNA. Quantified data that had been normalized to beta actin were converted into bar graph representations.

4. Multivariate Analysis of Prostate Disease State

a. Specimen Collection

Blood specimens (8-10 mls) were collected by venipuncture into standard serum or serum-separating tubes (Becton-Dickinson), allowed to coagulate for 30 minutes at room temperature, and then centrifuged at low speed (1000x *g*) for 10 minutes. Some specimens coming were immediately frozen and shipped overnight by delivery courier. Others were collected, processed, frozen, and shipped on dry ice by overnight mail. Upon arrival, all specimens were stored at -20°C. Repeated freeze-thaw cycles were avoided.

b. Measurement of Free and Total PSA

Two commercially available assays were utilized to measure PSA concentrations, an IMMULITE solid-phase chemiluminescence-based assay to measure free PSA (Diagnostic Products Corp.; Los Angeles, CA), and the FDA approved assay from TOSOH (San Diego, CA) that utilizes an enzyme-conjugated monoclonal antibody and fluorescent substrate to measure total PSA. However, since two different instruments were utilized to measure the components of the f/t PSA ratio, the international reference standards for free and total PSA were utilized to calibrate both assays and calculate the "corrected" f/t PSA ratio (Stamey, 1995).

c. f/t PSA Reference Standards and Correction of f/t PSA ratio

The corrected f/t PSA ratio was determined according to Marley *et al.*, 1996. Reference standards for free and total PSA assays were purchased from the Stanford University Prostate Center and consisted of an equimolar mixture of 90% PSA-α-1-antichymotrypsin and 10% free-PSA (Stamey, 1995; Chen *et al.*, 1995). All testing dilutions were performed with 1% bovine serum albumin (Fraction V; Sigma Chemical Co.) in 20 mM phosphate-buffered saline (PBS), pH 7.4. Expected concentrations of the reference standards, determined from molar extinction coefficients (ε), were also provided.

Free and total PSA assays were standardized as follows. Based upon the mean of seven linear standard curve runs of the reference standards (Stamey, 1995), correlation factors for free and total PSA measurement were calculated. Slope (m) deviations were measured relative to the linear plot based upon the PSA molar extinction coefficients (ε) of the reference standards. Since all curves passed through the origin, the correction factor for the free/total PSA ratio was calculated from the difference in slopes. Intra-assay coefficients of variation for free PSA (range = 0-2.0 ng/ml) and total PSA (range = 0-20.0 ng/ml) assays were 7% and 8%, respectively. The correction factors applied to the free and total PSA values were 1.19 and 0.83, respectively. For analysis purposes, only the f/t PSA ratio values were corrected.

The (TOSOH) total PSA assay reacted equally to the free and bound (PSA-ACT) forms of PSA. The (*Immulite*) free PSA assay system was unable to detect the bound fraction of PSA (PSA-ACT) below a concentration of 20 ng/ml. Antibodies for detecting both total and free PSA were unable to detect PSA covalently linked to α-2 macroglobulin (*PSA-MG* or occult PSA).

d. Statistical Methods

Differences in free and total serum PSA data between BPH and cancer samples were examined using the non-parametric statistical method of Wilcoxon rank-sum tests (Vollmer, 1996). The binary dependent variable assessed was the clinical outcome of BPH or CaP. Sensitivity, specificity and Receiver Operator Characteristics (ROC) Curves analyses were derived by Logistic regression modeling using the STATA™ software package (Stata Corporation, College Station, TX). Classification and Regression Tree (CART) analysis (CART v1.01, SYSTAT Inc., Evanston, IL), was used to determine the optimal cutoff for the serum assays as well as the logistic regression models (Breiman *et al.*, 1984; Steinberg and Colla, 1992). The correlation values of the independent parameters were also determined using the STATA™ software package.

e. IL-8 Quantitation

A commercial IL-8 immunoassay kit was purchased for use in this study (IL-8 Solid Phase Immunoassay, Cat. #D8050, 96 well microtiter plate format, from R&D Systems, 614 McKinley Pl. NE; Minneapolis, MN 55413). Solutions consisted of wash buffer, substrate solution (color reagents A&B), calibrator diluent RD6Z, assay diluent RD1-8, stop solution and IL-8 stock solution (2000 pg/ml). To prepare the IL-8 standards, 500 µl of calibrator diluent RD6Z was pipetted into each of a series of dilution tubes. A serial dilution of the IL-8 stock solution (2000 pg/ml) was prepared to yield standards of the following concentrations: 1000, 500, 250, 125, 62.5, 31.2, 15.6, 7.8 pg/ml.

The manufacturer's recommended protocol was used to assay IL-8 concentrations. All reagents and samples were first brought to room temperature. The assay mixture contained in each well; 100 µl of assay diluent RD1-8, 50 µl of sample (or appropriate standard) and 100 µl of IL-8 conjugate. The wells were covered with the provided adhesive strip and samples were incubated for 3 hours at room temperature. Each assay well was aspirated and washed with wash buffer for a total of six washes. After the final wash, the plate was inverted onto a paper towel to wick up excess moisture. Then 200 µl of substrate solution was added to each assay well and incubated for 30 min at room temperature. Fifty µl of stop solution was added to each assay well

and mixed by gentle tapping. Optical density was measured within 30 min of addition of stop solution, using a Bio-Tek EL-808 microplate reader (96 well format) at 450 nm.

f. IL-8 Standard Curve and Coefficient of Variation (CV)

The IL-8 standard curve consisted of eight concentrations: 1000, 500, 250, 125, 62.5, 31.2, 15.6, 7.8 pg IL-8/ml. The mean of six different measurements of each standard dilution was plotted (x-axis) vs. the mean optical density measured (y-axis). Results were plotted using the KC3 software package (Bio-Tek Instruments; Winooski, VT).

Coefficient of variation (CV): From the eight data points for each concentration of the standard curve, Coefficient of Variation (CV) = Standard Deviation/Mean was calculated to be 6.9, 6.4, 11.1, 10.1, 4.5, 4.4, 13.0 and 34.1%, respectively for the standard curve concentrations listed above. Points with a CV of greater than 13% were not utilized for this study.

K. EXAMPLES

Example 1

Relative Quantitative Reverse Transcriptase-PolymeraseChain Reaction - A method to evaluate novel genes (ESTs) as diagnostic biomarkers.

The reverse transcription-polymerase chain reaction (RT-PCR) protocols described in the following examples were developed as a means to determine the relative abundances of mRNA species that are expressed in various tissues, organs and cells. This protocol has been described as applied to prostate tissue in US Application Serial No. 08/692,787, incorporated in relevant part herein by reference. The protocols used to meet this need must be robust, reproducible, relatively quantitative, sensitive, conservative in its use of resources, rapid and have a high throughput rate. Relative quantitative RT-PCR has the technical features that, in theory, meet all of these criteria. In practice there are six important barriers to implementing an RT-PCR based assay that compares the relative abundances of mRNA species. The protocol described herein addresses each of these six barriers and has permitted the realization of the potential of RT-PCR for this application. Although the present example is drawn to the identification and confirmation of differential expression in various physiological states in prostate tissue, the methods described herein may be applied to any type of tissue, and particularly to peripheral blood cells to provide a sensitive method of identifying differential expression.

The inventors have described the examination of candidate genes by this method that were partial cDNA fragments identified by RNA fingerprinting methodologies. This necessitated development of a relatively quantitative approach to independently confirm the differential expression of the mRNAs from which these partial cDNA fragments were derived. The key objective of the described screening protocol is the assessment of changes in the relative abundances of mRNA.

The gene discovery program previously described is focused on analysis of human tissue and confirmation must be performed on the same biological material. Access to human tissue for isolation of RNA is limited. This limitation is especially problematic in Northern blots, the traditional means to determine differential gene expression. Northern blots typically consume roughly 20 μg of RNA per examined tissue per gene identified. This means that for the average size of tissue sample available, only 1-5 Northern blots can be performed before all of the RNA from a tissue sample is completely consumed. Clearly Northern blots are seriously limited for primary confirmation of discovered genes and consume extremely valuable biological resources required for gene discovery and characterization.

Because of such limitations on the amount of available tissue, and because of the need for high throughput and rapid turnaround of results, a two tiered assay protocol was developed that is technologically grounded on reverse transcription (RT) of RNA into cDNA followed by amplification of specific cDNA sequences by polymerase chain reaction (PCR). This coupling of techniques is frequently referred to as RT-PCR.

One advantage of RT-PCR is that it consumes relatively small quantities of RNA. With 20μg of RNA per examined sample, the amount of RNA required to perform a single Northern blot experiment, 50-200 RT-PCR assays may be performed with up to four data points per assay. Another advantage is a high throughput, eight independent experiments which examine eight different mRNA species for differential expression may be performed simultaneously in a single PCR machine with 96 wells. A single individual skilled in this technique may thereby examine and evaluate eight genes per day without significant time constraints. By comparison, even if RNA of sufficient quality and quantity were available to do this number of Northern blots, a similarly skilled individual performing Northern blots would be hard pressed to examine and evaluate eight genes per week. In addition to the lower throughput rate of Northern blots, eight

Northern blots per week would require the consumption of about 400μCi of ^{32}P per week. While not dangerous to use in the hands of a skilled individual, ^{32}P is certainly inconvenient to use. RT-PCR avoids the use of radioactive materials.

An additional advantage of RT-PCR over Northern blots as a technological platform for evaluating the relative expression of mRNA species is that RT-PCR is much less sensitive to differences in quality of the RNA being examined. The human tissues described were removed from patients for treatment purposes and were only incidentally saved for further studies. Hence the RNA, an extremely labile molecule, is expected to be at least partially degraded. Because the RNA is separated by size on a gel in the Northern blot assay, partially degraded RNA appears as a smear, rather than discrete bands. By contrast, RT-PCR amplifies only a section or domain of an RNA molecule, and as long as that portion is intact, the size or degradation state of the entire molecule is irrelevant. As a result, RNAs that are identical except that they vary by degree of partial degradation will give much more variable signals in a Northern blot than they will in an RT-PCR. When samples are of variable quality, as is often the case in human studies, the relative sensitivities of the techniques to variation in sample quality is an important consideration.

In the practice of this method, total cell RNA is first converted into cDNA using reverse transcriptase primed with random hexamers. This protocol results in a cDNA population in which each RNA has contributed according to its relative proportion in original total cell RNA. If two RNA species differ by ten fold in their original relative abundances in the total cell RNA, then the cDNA derived from these two RNAs will also differ by ten fold in their relative abundances in the resulting population of cDNA. This is a conservation of relative proportionality in the conversion of RNA to cDNA.

Another consideration is the relative rates of amplification of a targeted cDNA by PCR. In theory, the amount of an amplified product synthesized by PCR will be equal to $M(E^C)$. Where M is the mass of the targeted cDNA molecules before the beginning of PCR and C is the number of PCR cycle performed. E is an efficiency of amplification factor. This factor is complex and varies between 1 and 2. The important consideration in this assay is that over most of a PCR amplification, E will be nearly constant and nearly equal to 2. In PCR reactions that are identical in every way except the cDNAs being used as templates are derived from different total cell RNAs, then E will have the same value in each reaction. If a cDNA target has an initial mass

of M_1 in one PCR reaction and a mass of M_2 in another PCR reaction and if E has the same value in each reaction, then after C cycles of PCR there will be a mass of $M_1(E^C)$ of the amplified target in the first reaction and a mass of $M_2(E^C)$ of the amplified target in the second reaction. The ratios of these masses is unaltered by PCR amplification. That is $M1/M2= [M_1(E^C)]/M_2(E^C)$. Hence, there is a conservation of relative proportionality of amplified products during PCR.

Since both reverse transcription and PCR may be performed in such a way as to conserve proportionality, it is possible to compare the relative abundance of an mRNA species in two or more total cell RNA populations by first converting the RNA to cDNA and then amplifying a fragment of the cDNA derived from the specific mRNA by PCR. The ratio of the amplified masses of the targeted cDNA is very close to or identical to the ratios of the mRNAs in the original total cell RNA populations.

Six major challenges or barriers to be overcome in order to best use RT-PCR to quantify the relative abundances of RNA are as follows:

1) Degradation of RNA must be minimized during RNA preparation.
2) Genomic DNA must be eliminated.
3) RNA must be free of contaminants that might interfere with reverse transcription.
4) The efficiency of RT is variable. cDNAs, not RNA, must be normalized for equal concentrations of amplifiable cDNA.
5) Limited linear range requires multiple sampling points in any amplification curve.
6) Tube to tube variability in PCR

It is the development of techniques to overcome these barriers and to provide a sensitive and accurate method of quantitative RT-PCR that is applicable to any tissue type, or cell type such as peripheral blood cells, or physiological state that is a part of the present invention.

The first three barriers to successful RT-PCR are all related to the quality of the RNA used in this assay. The protocols described in this section address the first two barriers as described in the last section. These are the requirements that degradation of RNA must be minimized during RNA preparation and that genomic DNA must be eliminated from the RNA.

Two preferred methods for RNA isolation are the guanidinium thiocyanate method, which is well known in the art, and kits for RNA isolation manufactured by Qiagen, Inc. (Chatworth, CA), with the kits being the most preferred for convenience. Four protocols are performed on the RNA isolated by either method (or any method) before the RNA is be used in RT-PCR.

The first of these four protocols is digestion of the RNAs with DnaseI to remove all genomic DNA that was co-isolated with the total cell RNA. Prior to DNaseI digestion, the RNA is in a particulate suspension in 70% ethanol. Approximately 50 µg of RNA (as determined by $OD_{260/280}$) is removed from the suspension and precipitated. This RNA is resuspended in DEPC treated sterile water. To this is added 10X DNaseI buffer (200 mM Tris-HCl; pH 8.4, 20 mM $MgCl_2$, 500 mM KCl), 10 units of RNase Inhibitor (GIBCO-BRL Cat#15518-012) and 20 units of DNaseI (GIBCO-BRL # 18068-015). The volume is adjusted to 50 µl with additional DEPC treated water. The reaction is incubated at 37°C for 30 minutes. After DNaseI digestion the RNAs are organic solvent-extracted with phenol and chloroform followed by ethanol precipitation. This represents the second ethanol precipitation of the isolated RNA. Empirical observations suggest that this repeated precipitation improves RNA performance in the RT reaction to follow.

Following DNaseI digestion, an aliquot of the RNA suspension in ethanol is removed and divided into thirds. A different procedure is performed on each one of the aliquot thirds. These three procedures are: (1). An $OD_{260/280}$ is obtained using a standard protocol and is used to estimate the amount of RNA present and its likely quality. (2). An aliquot is run out on an agarose gel, and the RNA is stained with ethidium bromide. Observation that both the 28S and 18S RNAs are visible as discreet bands and that there is little staining above the point at which the 28S rRNA migrates indicate that the RNA is relatively intact. While it is not critical to assay performance that the examined RNAs be completely free of partial degradation, it is important to determine that the RNA is not so degraded as to significantly effect the appearance of the 28S rRNA. (3). The total cell RNAs are run using a PCR-based test that confirms that the DNaseI treatment actually digested the contaminating genomic DNA to completion. It is very important to confirm complete digestion of genomic DNA because genomic DNA may act as a template in PCR reactions resulting in false positive signals in the relative quantitative RT-PCR assay

described below. The assay for contaminating genomic DNA utilizes gene specific oligonucleotides that flank a 145 nucleotide long intron (intron #3) in the gene encoding Prostate Specific Antigen (PSA).This is a single copy gene with no pseudogenes. It is a member of the kallikrein gene family of serine proteases, but the oligonucleotides used in this assay are specific to PSA. The sequences of these oligonucleotides are:

5'CGCCTCAGGCTGGGGCAGCATT 3', SEQ ID NO:6

and

5'ACAGTGGAAGAGTCTCATTCGAGAT 3', SEQ ID NO:7.

In the assay for contaminating genomic DNA, 500 ng to 1.0 µg of each of the DNaseI treated RNAs are used as templates in a standard PCR (35-40 cycles under conditions described below) in which the oligonucleotides described above are used as primers. Human genomic DNA is used as the appropriate positive control. This DNA may be purchased from a commercial vender. A positive signal in this assay is the amplification of a 242 nucleotide genomic DNA specific PCR product from the RNA sample being tested as visualized on an ethidium bromide stained electrophoretic gel. There should be no evidence of genomic DNA as indicated by this assay in the RNAs used in the RT-PCR assay described below. Evidence of contaminating genomic DNA results in re-digestion of the RNA with DNaseI and reevaluation of the DNase treated RNA by determining its $OD_{260/280}$ ratio, examination on electrophoretic gel and re-testing for genomic DNA contamination using the described PCR assay.

The standard conditions used for PCR (as mentioned in the last paragraph) are:

1X GIBCO-BRL PCR reaction buffer [20 mM Tris-Cl (pH 8.4), 50 mM KCl]

1.5 mM $MgCl_2$

200 µM each of the four dNTPs

200 nM each oligonucleotide primer

concentration of template as appropriate

2.5 units of Taq polymerase per 100µl of reaction volume.

Using these conditions, PCR is performed with 35-40 cycles of:

94°C for 45 sec

55°-60°C for 45 sec

72°C for 1:00 minute.

The protocols described in the above section permit isolation of total cellular RNA that overcomes two of the six barriers to successful RT-PCR, *i.e.* the RNA is acceptably intact and is free from contaminating genomic DNA.

Reverse transcriptases, also called RNA dependent DNA polymerases, as applied in currently used molecular biology protocols, are known to be less processive than other commonly used nucleic acid polymerases. It has been observed that not only is the efficiency of conversion of RNA to cDNA relatively inefficient, there is also several fold variation in the efficiency of cDNA synthesis between reactions that use RNAs as templates that otherwise appear indistinguishable. The sources of this variation are not well characterized, but empirically, it has been observed that the efficiencies of some reverse transcription (RT) reactions may be improved by repeated organic extractions and ethanol precipitations. This implies that some of the variation in RT is due to contaminants in the RNA templates. In this case, the DNaseI treatment described above may be aiding the efficiency of RT by subjecting the RNA to an additional cycle of extraction with phenol and chloroform and ethanol precipitation. Contamination of the template RNA with inhibitors of RT is an important barrier to successful RT that is partially overcome by careful RNA preparation and repeated organic extractions and ethanol precipitations.

Reverse transcription reactions are performed using the Superscript™ Preamplification System for First Strand cDNA Synthesis kit which is manufactured by GIBCO-BRL LifeTechnologies (Gaithersburg, MD). Superscript™ is a cloned form of M-MLV reverse transcriptase that has been deleted for its endogenous Rnase H activity in order to enhance its processivity. In the present example, the published protocols of the manufacturer are used for cDNA synthesis primed with random hexamers. cDNA synthesis may also be primed with a mixture of random hexamers (or other small oligonucleotides of random sequence) and oligo dT. The addition of oligo dT increases the efficiency of conversion of RNA to cDNA proximal to the polyA tail. As template, either 5 or 10 micrograms of RNA is used (depending on availability). After the RT reaction has been completed according to the protocol provided by GIBCO-BRL, the RT reaction is diluted with water to a final volume of 100 µl.

Even with the best prepared RNA and the most processive enzyme, there may be significant variation in the efficiency of RT. This variation would be sufficiently great that

cDNA made in different RTs could not be reliably compared. To overcome this possible variation, cDNA populations made from different RT reactions may be normalized to contain equal concentrations of amplifiable cDNA synthesized from mRNAs that are known not to vary between the physiological states being examined. In the present examples, cDNAs made from total cell RNAs are normalized to contain equal concentrations of amplifiable b-actin cDNA.

One µl of each diluted RT reaction is subjected to PCR using oligonucleotides specific to β-actin as primers. These primers are designed to cross introns, permitting the differentiation of cDNA and genomic DNA. These β-actin specific oligonucleotides have the sequences:

5' CGAGCTGCCTGACGGCCAGGTCATC 3', SEQ ID NO:8

and

5' GAAGCATTTGCGGTGGACGATGGAG 3', SEQ ID NO:9

PCR is performed under standard conditions as described previously for either 19 or 20 cycles. The resulting PCR product is 415 nucleotides in length. The product is examined by PCR using agarose gel electrophoresis followed by staining with ethidium bromide. The amplified cDNA fragment is then visualized by irradiation with ultra violet light using a transilluminator. A white light image of the illuminated gel is captured by an IS-1000 Digital Imaging System manufactured by Alpha Innotech Corporation. The captured image is analyzed using either version 2.0 or 2.01 of the software package supplied by the manufacturer to determine the relative amounts of amplified β-actin cDNA in each RT reaction.

To normalize the various cDNAs, water is added to the most concentrated cDNAs as determined by the assay described in the last paragraph. PCR using 1 µl of the newly rediluted and adjusted cDNA is repeated using the β-actin oligonucleotides as primers. The number of cycles of PCR must be increased to 21 or 22 cycles in order to compensate for the decreased concentrations of the newly diluted cDNAs. With this empirical method the cDNAs may be adjusted by dilution to contain roughly equal concentrations of amplifiable cDNA. Sometimes this process must be repeated to give acceptable final normalization. By dividing the average optical density of all observed bands by that of a particular band, a normalization statistic may be created that will permit more accurate comparisons of the relative abundances of RNAs examined in the normalized panel of cDNAs.

Once the normalization statistics are derived, PCR may be performed using different gene specific oligonucleotides as primers to determine the relative abundances of other mRNAs as represented as cDNAs in the normalized panel of diluted RT reaction products. The relative intensities of the bands is then adjusted and normalized to β-actin expression by multiplying the intensity quantities by the normalization statistics derived.

In the next section an RT-PCR assay is discussed that uses pooled cDNAs and is more likely to capture data from PCRs while in the linear portions of their amplification curves. The error caused by observing PCRs after the linear portion of PCR is in the direction of quantitatively underestimating mRNA abundance differences. To determine quantitative differences in mRNA expression, it is necessary that the data is collected in the linear portion of the respective PCR amplification curves. This requirement is met in the assay described in following paragraphs.

The last two barriers to RT-PCR are addressed in the sections that follow involving the use of pooled cDNAs as templates in RT-PCR. In practice, the protocols using pooled templates are usually performed before the protocol described above.

There are two additional barriers to relative mRNA quantitation with RT-PCR that frequently compromise interpretations of results obtained by this method. The first of these involves the need to quantify the amplification products while the PCR is still in the linear portion of the process where "E" behaves as a constant and is nearly equal to two. In the "linear" portion of the amplification curve, the log of the mass of the amplified product is directly proportional to the cycle number. At the end of the PCR process, "E" is not constant. Late in PCR, "E" declines with each additional cycle until there is no increase in PCR product mass with additional cycles.

The most important reason why the efficiency of amplification decreases at high PCR cycle number, may be that the concentration of the PCR products becomes high enough that the two strands of the product begin to anneal to each other with a greater efficiency than that at which the oligonucleotide primers anneal to the individual product strands. This competition between the PCR product strands and the oligonucleotide primers creates a decrease in PCR efficiency. This part of the PCR where the efficiency of amplification is decreased is called the "plateau" phase of the amplification curve. When "E" ceases to behave as a constant and the

PCR begins to move towards the plateau phase, the conservation of relative proportionality of amplified products during PCR is lost. This creates an error in estimating the differences in relative abundance of an mRNA species occurring in different total cell RNA populations. This error is always in the same direction, in that it causes differences in relative mRNA abundances to appear less than they actually are. In the extreme case, where all PCRs have entered the plateau phase, this effect will cause differentially expressed mRNAs to appear as if they are not differentially expressed at all.

To control for this type of error, it is important that the PCR products be quantified in the linear portion of the amplification curve. This is technically difficult because currently used means of DNA quantitation are only sensitive enough to quantify the PCR products when they are approaching concentrations at which the product strands begin to compete with the primers for annealing. This means that the PCR products may only be detected at the very end of the linear range of the amplification curve. Predicting in advance at what cycle number the PCR products should be quantified is technically difficult.

Practically speaking, it is necessary to sample the PCR products at a variety of cycle numbers that are believed to span the optimum detection range in which the products are abundant enough to detect, but still in the linear range of the amplification curve. It is impractical to do this in a study that involves large numbers of samples because the number of different PCR reactions and/or number of different electrophoretic gels that must be run becomes prohibitively large.

To overcome these limitations, a two tiered approach was designed to relatively quantify mRNA abundance levels using RT-PCR. In the first tier, pools of cDNAs produced by combining equal amounts of normalized cDNA are examined to determine how mRNA abundances vary in the average individual with a particular physiological state. This reduces the number of compared samples to a very small number such as two to four. In the studies described herein, two pools are examined. These are pools of normal individuals and those individuals with metastatic prostate cancer. Each pool may contain a large number of individuals. While this approach does not discriminate differences between individuals, it may easily discern broad patterns of differential expression. The great advantage of examining pooled cDNAs is that it permits many duplicate PCR reactions to be simultaneously set up.

The individual duplicates may be harvested and examined at different cycle numbers of PCR. In studies described below, four duplicate PCR reactions were set up. One duplicate was collected at 31, 34, 37, and 40 PCR cycles. Occasionally, PCR reactions were also collected at 28 cycles. Examining the PCRs at different cycle numbers yielded the following benefits. It is very likely that at least one of the RT-PCRs will be in the optimum portion of the amplification curves to reliably compare relative mRNA abundances. In addition, the optimum cycle number will be known, so that studies with much larger sample sizes are much more likely to succeed. This is the second tier of a two tiered approach that has been taken to relatively quantify mRNA abundance levels using RT-PCR. Doing the RT-PCR with the pooled samples permits much more efficient application of RT-PCR to the samples derived from individuals. A further benefit, also as discussed below, tube to tube variability in PCR may be discounted and controlled because most studies yield multiple data points due to duplication.

Like the previously described protocol involving individuals, the first step in this protocol is to normalize the pooled samples to contain equal amounts of amplifiable cDNA. This is done using oligonucleotides that direct the amplification of β-actin. In this example, a PCR amplification of a cDNA fragment derived from the β-actin mRNA from pools of normal individuals and individuals with metastatic prostate cancer was performed. This study was set up as four identical PCR reactions. The products of these PCRs were collected and electrophoresed after 22, 25, 28 and 31 PCR cycles. Quantitation of these bands using the IS 1000 system showed that the PCRs were still in the linear ranges of their amplification curves at 22, 25 and 28 cycles but that they left linearity at 31 cycles. This is known because the ratios of the band intensities remain constant and internally consistent for the data obtained from 22, 25 and 28 cycles, but these ratios become distorted at 31 cycles. This quantitation will also permit the derivation of normalizing statistics for the three pools relative to each other in exactly the same manner as was done previously for individuals.

This study is then repeated using gene specific primers for a gene other than β-actin. The intensities of the relevant bands were quantitated using the IS 1000 and normalized to the β-actin signals.

The central question to be answered in analyzing this data is whether the PCRs have been examined in the linear portions of their amplification curves. A test for this may be devised by

determining if the proportionality of the PCR products has been conserved as PCR cycle number has increased. If the ratio between the two pools of a given PCR product remains constant with increasing cycle number, this is strong evidence that the PCRs were in the linear portions of their amplification curves when these observations were made. (This is better conservation of proportionality than is frequently observed. In some studies, data was excepted when the ratios were similar but not identical.) This conservation of proportionality was lost at 40 cycles. This indicates that these PCRs are nearing the plateau phases of their amplification curves.

The final major barrier to quantifying relative mRNA abundances with RT-PCR is tube to tube variability in PCR. This may result from many factors, including unequal heating and cooling in the thermocycler, imperfections in the PCR tubes and operator error. To control for this source of variation, the Cole-Parmer digital thermocouple Model # 8402-00 was used to calibrate the thermocyclers used in these studies. Only slight variations in temperature were observed.

To rigorously demonstrate that PCR tube to tube variability was not a factor in the studies described above, 24 duplicate PCRs for β-actin using the same cDNA as template were performed. These PCR tubes were scattered over the surface of a 96 well thermocycler, including the corners of the block where it might be suspected the temperature might deviate from other areas. Tubes were collected at various cycle numbers. Nine tubes were collected at 21 cycles. Nine tubes were collected at 24 cycles, and six tubes were collected at 27 cycles. Quantitation of the intensities of the resulting bands with the IS 1000 system determined that the standard error of the mean of the PCR product abundances was ±13%. This is an acceptably small number to be discounted as a major source of variability in an RT-PCR assay.

The RT-PCR protocol examining pooled cDNAs is internally controlled for tube to tube variability that might arise from any source. By examining the abundance of the PCR products at several different cycle numbers, it may be determined that the mass of the expected PCR product is increasing appropriately with increasing PCR cycle number. Not only does this demonstrate that the PCRs are being examined in the linear phase of the PCR, where the data is most reliable, it demonstrates that each reaction with the same template is consistent with the data from the surrounding cycle numbers. If there was an unexplained source of variation, the expectation that PCR product mass would increase appropriately with increasing cycle number would not be met.

This would indicate artifactual variation in results. Internal duplication and consistency of the data derived from different cycle numbers controls for system derived variation in tube to tube results.

As described in the preceding paragraphs, the RT-PCR protocol using pooled cDNA templates overcomes the last two barriers to effective relative quantitative RT-PCR. These barriers are the need examine the PCR products while the reactions are in the linear portions of their amplification curves and the need to control tube to tube variation in PCR. The described protocol examines PCR products at three to four different cycle numbers. This insures that the PCRs are quantitated in their linear ranges and, as discussed in the last paragraph, controls for possible tube to tube variation.

One final question is whether β-actin is an appropriate internal standard for mRNA quantitation. β-actin has been used by many investigators to normalize mRNA levels. Others have argued that β-actin is itself differentially regulated and therefore unsuitable as an internal normalization standard. In the protocols described herein differential regulation of β-actin is not a concern. More than fifty genes have been examined for differential expression using these protocols. Fewer than half were actually differentially expressed. The other half were regulated similarly to β-actin within the standard error of 13%. Either all of these genes are coordinately differentially regulated with β-actin, or none of them are differentially regulated. The possibility that all of these genes could be similarly and coordinately differentially regulated with β-actin seems highly unlikely. This possibility has been discounted.

β-actin has also been criticized by some as an internal standard in PCRs because of the large number of pseudogenes of β-actin that occur in mammalian genomes. This is not a consideration in the described assays because all of the RNAs used herein are demonstrated to be free of contaminating genomic DNA by a very sensitive PCR based assay. In addition, the cycle number of PCR needed to detect β-actin cDNA from the diluted RT reactions, usually between 19 and 22 cycles, is sufficiently low to discount any contribution that genomic DNA might make to the abundance of amplifiable β-actin templates.

Example 2:

Identification of Markers of Metastatic Prostate Cancer by Use of RNA fingerprinting by PCR primed with oligonucleotides of arbitrary sequence.

RNA fingerprinting by PCR, primed with oligonucleotides of arbitrary sequence was performed on RNAs isolated from peripheral human blood. Bands which appeared to be differentially expressed were cloned.

For this study, total cell RNA was isolated from buffy coat cells as described above. cDNA was made from one to five μg of each isolated RNA. All cDNAs were normalized for similar amounts of ß-actin cDNA by RT-PCR. RT-PCR products were electrophoresed through agarose.

For relative quantitative RT-PCR with an external standard, quantitation of band intensities on ethidium bromide stained gels was performed using the IS-1000 image analysis system manufactured by the Alpha Innotech Corp. A normalizing statistic was generated for each cDNA sample, as the average of all ß-actin signals divided by the ß-actin signal for each cDNA sample respectively. Data for each sample was then normalized by multiplying the observed densitometry observation by the individual normalizing statistics. Normalized values predict differences in the steady state abundances of the respective mRNAs in the original total cell RNA samples.

The nucleotide sequences of all cloned PCR products were determined by dideoxy termination sequencing using either the ABI or Pharmacia automated sequencers.

This protocol resulted in the discovery of an mRNA species that was 2-3 fold less abundant in the peripheral blood of metastatic prostate cancer patients than in the peripheral blood of normal individuals of both sexes. The sequence of this band, referred to as UCBP Band #35 (SEQ ID NO:1), matches an EST derived from a fetal brain cDNA library (GenBank Accession #T03013). Down regulation of this band in the peripheral blood of metastatic prostate cancer patients was confirmed by relative quantitative RT-PCR.

Example 3:

Identification of Markers of Metastatic Prostate Cancer by Use of RNA fingerprinting by the Pairwise Sequential Method.

RNA fingerprinting was used to identify differentially expressed RNA species according to the pairwise sequential method of McClelland *et al.* (1994), as modified to use larger (17-25 mer)

arbitrary oligonucleotides. PCR amplification products were labeled using α-32P-dCTP and were visualized by autoradiography after electrophoresis on denaturing polyacrylamide gels. A number of bands appeared to be differentially expressed, and were cloned as described above.

UC Band #321 was confirmed by RT-PCR to be down regulated in the peripheral blood of prostate cancer patients, with a four-fold decrease observed compared with normal individuals. The DNA sequence of Band #321 does not match any known sequences in the GenBank database. It therefore represents a previously undescribed gene product.

UC Band #302 and UC Band #325 were both observed to be up regulated in the peripheral blood of metastatic prostate cancer patients. UC Band #302 is identical in sequence to a portion of the sequence of elongation factor 1-α (GenBank Accession #X03558). This band was modestly increased between 1.6 and 2-fold in metastatic cancer patients compared with normal individuals.

UC Band #325 was found to consist of two different alternatively spliced forms of mRNA, encoded by the interleukin-8 (IL-8) gene. UC Band #325-1, the previously identified mRNA species of IL-8 (Genbank Accession #Y00787), is approximately seven-fold more abundant in the peripheral blood of metastatic prostate cancer patients. The alternatively spliced IL-8 mRNA, containing intron #3 of the IL-8 gene (Genbank Accession #M28130) is up to seven-fold less abundant in the peripheral blood of metastatic prostate cancer patients. Fig. 1A shows relative quantitative RT-PCR of the differential expression of IL-8 (=UC235) in peripheral blood of patients with metastatic prostate cancer (M) and normal individuals (N) at different PCR cycles (cy). The two alternatively spliced forms of the IL-8 mRNA are observed. The upper band (int.+) includes intron 3 in the mature mRNA. The lower band (int.-) lacks intron 3. Fig. 1B shows relative quantitative RT-PCR showing Differential Expression of IL-8 (UC325) in peripheral blood of patients with metastatic prostate cancer in lanes 1-5 and a pool of normal individuals (N). The alternatively spliced forms of the IL-8 mRNA observed are different between normal individuals and those with prostate cancer. Overall, there is an approximately 30-fold change in the ratios of the two spliced forms of IL-8 mRNA in individuals with metastatic prostate cancer compared with normal individuals. These results have been confirmed by relative quantitative RT-PCR.

As described above, an increased expression of IL-8 mRNA has been previously reported in cancer patients. However, this represents the first finding of an alternatively spliced form of IL-8 mRNA, containing intron 3, that is significantly more abundant in normal individuals compared

with metastatic prostate cancer patients. These results are surprising in view of previous reports which had failed to find any alternatively spliced forms of IL-8 mRNA in normal individuals or cancer patients.

It will be recognized that the genes and gene products (RNAs and proteins) for the above described markers of metastatic prostate cancer are included within the scope of the disclosure herein described. It will also be recognized that the diagnosis and prognosis of metastatic prostatic cancer by detection of the nucleic acid products of these genes are included within the scope of the present invention. Serological and other assays to detect these mRNA species or their translation products are also indicated. It is obvious that these assays are of utility in diagnosing metastatic cancers derived from prostate and other tissues.

Most significantly, these Examples demonstrate the feasibility of using RNA fingerprinting to identify mRNA species that are differentially expressed in the peripheral blood of patients with asymptomatic diseases or in patients with symptoms that are insufficient for a definitive diagnosis. It will be appreciated that this technique is applicable not only to the detection and diagnosis of prostate and other cancers, but also to any other disease states which produce significant effects on lymphocyte gene expression. Uses which are contemplated within the scope of the present disclosure include the detection and diagnosis of clinically significant diseases that requires medical intervention, including but not limited to asthma, lupus erythromatosis, rheumatoid arthritis, multiple sclerosis, myasthenia gravis, autoimmune thyroiditis, ALS, interstitial cystitis and prostatitis.

TABLE 2

Genes Whose mRNAs have Abundances that Vary in Metastatic Prostate Cancer Relative to Normal Individuals

Name of cDNA Fragment	Sequence Determined	Confirmed by RT-PCR	Previously Known
UCPB 35	Yes	Yes	GB #T03013
UC 302 SEQ ID NO:3	Yes	Yes	EF 1-α
UC 321 SEQ ID NO:2	Yes	Yes	No
UC 325-1 SEQ ID NO:4	Yes	Yes	GB #Y00787
UC 325-2 SEQ ID NO:5	Yes	Yes	IL-8

TABLE 3.

Oligonucleotides used in the relative quantitative RT-PCR portion of these studies.

Oligonucleotides used to examine the expression of genes:

UCPB Band #35 (previously uncharacterized gene).

5' TGCAAACTTTCACCTGGACTT3', SEQ ID NO:10

5' CTTGTGACTTGCTTTGATAGAATG3', SEQ ID NO:11

UC Band #302 (elongation factor 1-α).

5' GACAACATGCTGGAGCCAAGTGC3', SEQ ID NO:12

5' ACCACCAATTTTGTAAGAACATCCT3', SEQ ID NO:13

UC Band #321 (previously uncharacterized gene).

5' TGTCCAGAGATCCAAGTGCAGAAGG3', SEQ ID NO:14

5' GAGCTCCAGGAGACAGAAGCCATAG3', SEQ ID NO:15

UC Band #325-1 (IL-8).

5' GGGCCCCAAGGAAAACT3', SEQ ID NO:16

5' TGGCAACCCTACAACAGACC3', SEQ ID NO:17

UC Band #325-2 (IL-8).

5' GGGCCCCAAGGAAAACT3', SEQ ID NO:18

5' TGGCAACCCTACAACAGACC3', SEQ ID NO:19

Controls used to normalize relative quantitative RT-PCR

β-actin

5' CGAGCTGCCTGACGGCCAGGTCATC3', SEQ ID NO:8

5' GAAGCATTTGCGGTGGACGATGGAG3', SEQ ID NO:9

Asparagine Synthetase (AS)

5' ACATTGAAGCACTCCGCGAC3', SEQ ID NO:20

5' AGAGTGGCAGCAACCAAGCT3', SEQ ID NO:21

Example 4:

DNA Sequences of Markers of Metastatic Prostate Cancer

The DNA sequences of the markers of metastatic prostate cancer were determined by Sanger dideoxy sequencing as detailed above. The identified sequences are provided in Table 4.

TABLE 4.

DNA Sequences of Markers of Metastatic Prostate Cancer:

UCPB Band #35 (SEQ ID NO:1) Matches a fetal brain EST, GenBank Accession # T03013

5'GGCAGGGGCTTGTGACTCTAAGATGGCTTCATTCACATGCCTAGGGCCTCAGTAGG
ATGACTGGCATGGCCCTGGAAAACTGCGAAGTCTTCTCTCTGTGCAAACTTTCACCT
GGACTTTTTATATGATTCTGGAAGTATTCCAAGAAGGCAAAAGTAAAAACTGCAAA
GCGTCTTAAAATAGAAGTTCAGAAGCCACATTATATCACTTCTGTTGCATTCTATCA
AAGCAAGTCACAAGCCCCTGCCAATCA 3'

UC Band # 321 (SEQ ID NO:2) previously uncharacterized Gene

5'CACACACTCCCCCATTCTGAGCCCCAAGAGGCTCATCCCTAAGGATGTCCAGAGA
TCCAAGTGCAGAAGGAGAATGTGGTGAGGCTATTTATTCCCCCAGTGCCTTCCCTGC
TGGGCTATGGATGAACAGTGGCTGACTTCATCTAGGAAAGAGCTATGGCTTCTGTCT
CCTGGAGCTCACCA 3'

UC Band # 302 (SEQ ID NO:3) Human Elongation Factor 1-alpha, GenBank Accession #X03558

5'GGTGAGCCCCAGGAGACAGAAGAGATATGAGGAAATTGTTAAGGAAGTCAGCAC
TTACATTAAGAAAATTGGCTACAACCCCGACACAGTAGCATTTGTGCCAATTTCTGG
TTGGAATGGTGACAACATGCTGGAGCCAAGTGCTAACATGCCTTGGTTCAAGGGAT
GGAAAGTCACCCGTAAGGATGGCAATGCCAGTGGAACCACGCTGCTTGAGGCTCTG
GACTGCATCCTACCACCAACTCGTCCAACTGACAAGCCCTTGCGCCTGCCTCTCCAA
GGATGTTCTTACAAAATTGGTGGTATTGGTACTGTTCCCTGTTTGGCCGAATTGGAA
AACTGGTGTTCCTCCAAACCCCGGTTATGGTGGGTTTCCTCCTCCTTGGA 3'

UC Band #325-1 (SEQ ID NO:4) Human IL-8 mRNA, GenBank Accession #Y00787

5'GGGCGGAACAAGGGAGCGCTAAAAGGAAATTAGGATGTCAGGTGCATAAAGGAC
ATAATTCCAAAACCTTTCCAAACCCCAAATTTATTCAAAGGAACTGAGGAGTGGATT
GAGGAGTGGACCAACACTGGCGCCAAACACAGAAATTATTGTAAAGCTTTCTGATG

GAAGAGAGCTCTGTCTGGGCCCCAAGGAAAACTGGGTGCAGAGGGTTGTGGAGAAG
TTTTTGAAGAGGGCTGAGAATTCATAAAAAAATTCATTCTCTGTGGTATCCAAGAAT
CAGTGAAGATGCCAGTGAAACTTCAAGCAAATCTACTTCAACACTTCATGTATTGTG
TGGGTCTGTTGTAGGGTTGCCAGTTGTT 3'

UC Band #325-2 (SEQ ID NO:5) Human IL-8 mRNA Containing Intron #3.

5'GCTTGGGCCCCAAGGAAAACTGGGTGCAGAGGGTTGTGGAGAAGTTTTTGAAGAG
GTAAGTTATATATTTTTGAATTTAAAATTTGTCATTTATCCGTGAGACATATAATCCA
AAGTCAGCCTATAAATTTCTTTCTGTTGCTAAAAATCGTCATTAGGTATCTGCCTTTT
TGGTTAAAAAAAAAAGGAATAGCATCAATAGTGAGTGTGTTGTACTCATGACCAGA
AAGACCATACATAGTTTGCCCAGGAAATTCTGGGTTTAAGCTTGTGTCCTATACTCTT
AGTAAAGTTCTTTGTCACTCCCAGTAGTGTCCTATGTTAGATGATAATGTCTTTGATC
TCCCTATTTATAGTTGAGAATATAGAGCATGTCTAACACATGAATGTCAAAGACTAT
ATTGACTTTTCAAGAACCCTACTTTCCTTCTTATTAAACATAGCTCATCTTTATATTGT
GAATTTTATTTTAGGGCTGAGAATTCATAAAAAAATTCATTCTCTGTGGTATCCAAG
AATCAGTGAAGATGCCAGTGAAACTTCAAGCAAATCTACTTCAACACTTCATGTATT
GTGTGGGTCTGTTGTAGGGTTGCCA 3'

Example 5:

Detection and Differential Diagnosis of BPH versus Localized and Advanced Stage Prostate Carcinomas Using Combinations of IL-8 with Other Prostate Disease Markers.

A total of 164 serum specimens from normal men or men with a biopsy confirmed diagnosis of BPH or prostate cancer were studied. These serum specimens were provided by Dr. George Wright from the Virginia Prostate Center at the Eastern Virginia Medical School and by Dr. Robert Vessella from the University of Washington or were normal donors from UroCor, Inc. All patients were biopsy-confirmed for either BPH or prostate carcinoma (stages A, B, and C only) within six months after PSA serum collection and/or a DRE-positive diagnosis. All patient sera were obtained prior to any surgical or hormonal therapies. The mean age of the total sample was 69.4 ± 8.6 years (range = 37 - 91 years) old.

The subset of patients utilized for multivariate diagnostic serum model consisted of 13 BPH and 64 CaP (Stages A, B, C) cases from the parent population (Marley *et al.*, 1996). All patients in the subset had a total PSA between 2.0 - 20.0 ng/ml, which is a standard range for f/t

PSA testing (Marley *et al.*, 1996). Also evaluated were a subset of Stage D CaP patients, with t-PSA values ranging from 6.5 - 867 ng/ml.

Diagnosis	N	Mean Age ± Std. Dev. (Range)
Normal	8	< 50 years
BPH	55	66.4± 8.6 (37 - 87) years
CaP Stage A	24	74.7±7.8 (61 - 91) years
CaP Stage B	48	68.3±7.9 (51 - 85) years
CaP Stage C	14	68.9±6.9 (60 - 80) years
CaP Stage D	14	72.3±8.6 (58 - 86) years

Table 5 shows the distribution of the total PSA levels, the f/t PSA ratios, and the UC325 levels for the 164 patients, broken down by normals, BPH, and Stages A, B, C, & D prostate cancer. Only the BPH, Stage A, Stage B, and Stage C prostate cancer patients were included in the statistical analysis.

TABLE 5

UC325 Patient Sample Characteristics (n = 164)

		Mean Value ± Std. Dev.		
Diagnosis	N	UC325 (pg/ml)	Total PSA (ng/ml)	f/t PSA Ratio (%)
Normal	8	0.2 ± 0.6	N/A	N/A
BPH	55	6.8 ± 6.1	6.9 ± 4.0	21.9 ± 10.9%
CaP Stage A	24	19.1 ± 10.4	6.2 ±2.7	14.6 ± 10.5%
CaP Stage B	48	13.5 ± 9.5	8.8 ± 6.6	11.9 ± 5.7%
CaP Stage C	15	19.1 ± 7.9	16.2 ± 7.6	11.2 ± 8.3
CaP State D	14	78.9 ± 197	244 ± 332	12.4 ± 7.1%

Table 6 illustrates the ability for f/t PSA ratio at three different cutoffs to differentiate prostate cancer and BPH in the inventors' patient sample. UC325 (IL-8) and t-PSA are analyzed at single Classification and Regression Tree (CART) cutoff points for the same outcome. Note the significant improvement in both sensitivity and specificity contributed by the UC325 (IL-8) serum assay to detect clinically organ confined. The combination of UC325 (IL-8), treated as a continuous variable, and t-PSA or f/t PSA ratio provides a highly predictive multivariate test

system to diagnose CaP (clinical stages A & B) without any interference provided by BPH in the inventors' patient subset.

TABLE 6

Ability of Serum Tests to Discriminate BPH and CaP.

Serum Test	Cutoff	Sensitivity	Specificity	AUC	p-value
f/t PSA Ratio	11%	52.9%	91.9%	0.7905	<0.0001
""	14%	70.1%	80.0	""	""
""	20%	85.1	47.3	""	""
UC325	9.8 pg/ml	72.4%	74.5%	0.7973	<0.0001
Total PSA	14.8 ng/ml	17.2%	98.2%	0.5995	0.0134
f/t PSA & UC325	0.69**	71.3%	90.9%	0.8784	<0.0001
Total PSA & UC325	0.64**	62.1%	85.5%	0.8069	<0.0001

*All cutoff's determined using Glassification and Regression Tree Analysis (CART)

**Predicated Probability value calculated using logistic regression function

To further substantiate the results of Table 6, individual analysis using Receiver Operator Characteristic (ROC) curves are provided for each variable. Figure 2 illustrates the ability of t-PSA to distinguish BPH and Stages A, B, and C prostate cancer. Figure 3 shows the ability of f/t PSA ratio to distinguish BPH and Stages A, B, and C prostate cancer. Figure 4 shows the ability of UC325 (IL-8) alone to distinguish BPH and Stages A, B, and C prostate cancer. Figure 5 shows the ability of the combination of UC325 (IL-8) and total PSA (t-PSA) to distinguish BPH and Stages A, B and C prostate cancer. Figure 6 shows the ability of the combination of UC325 (IL-8) and the f/t PSA ratio to distinguish between BPH and stages A, B and C prostate cancer. It is apparent that the combination of UC325 measurement with either t-PSA or f/T PSA provides a significant increase in sensitivity of detection, while maintaining a high degree of specificity. Thus, the combination of UC325 (IL-8) with other prostate disease markers, such as t-PSA or f/t PSA ratio, provides a significant advance in the detection and differential diagnosis of prostate cancer.

Table 7 presents the correlation values for the different serum markers. This table clearly shows that the UC325 biomarker provides information which is independent of that provided by the f/t PSA ratio.

TABLE 7

Correlation Values for BPH vs Stages A, B & C (n = 142)

	Diagnosis	Total PSA (ng/ml)	f/t PSA Ratio (%)	UC325 (pg/ml)	Age	Clinical Stage
Diagnosis	1.0000	0.5647	-0.1912	0.2262	0.1590	0.3497
Total PSA (ng/ml)	0.5647	1.000	-0.2319	0.5991	0.0898	0.3729
f/t PSA Ratio (%)	-0.1912	0.2319	1.0000	-0.2142	0.0641	-0.4126
UC325 (pg/ml)	0.2262	0.5991	0.2142	1.0000	0.0881	0.2486
Age	0.1590	0.0898	0.0641	0.0881	1.0000	0.1372
Clinical Stage	0.3497	0.3729	-0.4126	0.2486	0.1372	1.0000

Table 8 clearly demonstrates a relationship between tumor burden and serum UC-325 gene product measured by IL-8 assay. Note that as biopsy-confirmed clinical stage of the cancer increases, so does the IL-8 serum marker concentration, whereas the same relationship did not occur with [t-PSA] or f/t PSA ratio.

TABLE 8

UC325 Culled Dataset, One High and Low Value Removed (n=164)

Specimen Stage	N	UC325 (10 pm/ml Cutoff) Negative	UC325 (10 pm/ml Cutoff) Positive	UC325 (15 pg/ml Cutoff) Negative	UC325 (15 pg/ml Cutoff) Positive
Normal	8	8 (100%)	0 (0%)	8 (100%)	0 (0%)
BPH	55	41 (75%)	14 (25%)	50 (91%)	5 (9%)
Stage A & B	72	25 (35%)	47 (65%)	43 (60%)	29 (40%)
Stage C	15	0 (0%)	15 (100%)	5 (33%)	10 (67%)
Stage D	14	2 (14%)	12 (86%)	3 (21%)	11 (79%)

Example 6:

Identification of Markers of Metastatic Prostate and Breast Cancer by Use of RNA fingerprinting by PCR primed with oligonucleotides of arbitrary sequence.

RNA fingerprinting displays PCR™ amplified cDNA fragments that represent a sample of RNA species derived from a population of total cell RNAs. When displayed side by side, comparisons of similarly produced fingerprints representing RNA populations from cells of differing physiologic states identifies mRNA species whose relative abundances vary between the examined physiologic states. In this study, RNA fingerprinting identified two cDNA fragments derived from mRNA species that had higher steady state abundances in the peripheral blood leukocytes of patients with recurrent metastatic prostate cancer as compared to a group of healthy volunteers.

Eight ml of peripheral blood was collected from healthy volunteers, patients with clinically and biopsy confirmed BPH, localized and advanced metastatic prostate cancer, and from patients with advanced metastatic breast cancer. Metastatic prostate and breast cancer patients that had failed a primary therapy and had evidence of recurrence of disease were selected. The metastatic prostate cancer patients had high (≥ 50 ng/ml) serum concentrations of PSA. Circulating nucleated peripheral blood cells were separated from erythrocytes by centrifugation in Vacutainer© CPT™ tubes (Becton Dickinson and Company, Franklin Lakes, N J). Total RNA was prepared from isolated nucleated peripheral blood cells by lysis with RNA Stat-60™ (Tel-Test, Inc., Friendswood, TX) following the instructions provided by the vendor. Contaminating genomic DNA was removed from the total RNAs by digestion with RNase free DNaseI (GIBCO-BRL, Gaithersburg, MD). For the PCR™ based applications of RNA fingerprinting and relative quantitative RT-PCR™, it is absolutely critical that the total RNA is completely free of genomic DNA. Typically, 5.0 to 10.0 μg of total RNA was digested with 20-40 units of RNase free DNaseI in 100-200 μl of reaction volume for 20 min at 37°C.

Following digestion, the total RNAs were extracted with phenol (pH=4.3, Amresco, Inc., Solon, OH) and ethanol precipitated. To confirm that the RNA was free of contaminating genomic DNA, 500 ng to 1.0 μg of each DNaseI treated RNA was resuspended in water. These were used as templates for PCR™ using oligonucleotide primers that anneal to exons 3 and 4 of

the gene encoding PSA (exon 3: 5′ GCCTCAGGCTGGGGCAGCATT 3′ SEQ ID NO:22, exon 4: 5′ GGTCACCTTCTGAGGGTGAACTTGC 3′ SEQ ID NO:23). These primers anneal to opposite strands of genomic DNA that flank the 145 bp intron 3 of the PSA gene. PCR™ was performed at 94°C for 1:15 min, followed by 40 cycles of 94°C for 45 sec, 55°C for 45 sec, and 72°C for 1:15 min, then a final extension of 72°C for 5:00 min. RNA was considered DNA-free if no PCR™ products could be visualized upon gel electrophoresis that co-migrated with the PSA gene positive control of known human genomic DNA. If PSA gene products were observed after PCR™, the RNA was redigested with DNaseI and analyzed again for contaminating genomic DNA. After it was confirmed that the RNAs were free of genomic DNA, 500 ng to 1.0 μg of RNA was electrophoresed on a 1.2% agarose Tris Acetate EDTA (TAE) to visualize the ribosomal RNAs (Fridell *et al.*, 1995). Only RNA preparations for which the 28S ribosomal RNA could be visualized were selected for further analysis by RNA fingerprinting and relative quantitative RT-PCR™.

RNA fingerprinting with arbitrarily chosen oligonucleotide primers (Welsh *et al.*, 1992) is conceptually similar to differential display (Liang and Pardee, 1992), except that oligonucleotides of arbitrary sequence are used to prime both strands of cDNA synthesis instead of just second strand synthesis, as in differential display. In this investigation, the strategy of RNA fingerprinting used was similar to that described in Ralph *et al.* (1993) except that oligonucleotide primers used were composed of two discrete domains. The 5′ domain of these oligonucleotides consisted of ten nucleotides that complemented sequences from either the T7 promotor or the M13 reverse sequencing primer. The 3′ domains of these oligonucleotides were 8-mer sequences predicted to anneal frequently to the protein-coding regions of mRNAs in a permiscuous fashion (Lopez-Nieto and Nigam, 1996). These oligonucleotides were then used in a sequential pairwise strategy that optimizes the amount of mRNA complexity that can be surveyed with limited numbers of primers and starting RNA. Care was taken to ensure that the two oligonucleotides used to produce any single fingerprint did not share sequence similarity in either their 5′ or 3′ domains. Because these oligonucleotides were constructed of short sequence domains that have specific functions within this experimental design, the oligonucleotides are permiscuous rather than truly arbitrary in nature.

Two RNA pools were fingerprinted. These two pools were each created by combining equal amounts of peripheral blood total RNA from five individuals. One pool was constructed by pooling RNA from five healthy individuals while the other pool was derived from five individuals with recurring metastatic prostate cancer. Using the pooled RNAs as templates, first strand cDNA synthesis was primed by annealing one of the permiscuous oligonucleotide primers to the pooled RNAs at low stringency. All fingerprinting studies were performed in duplicate using different initial concentrations of template RNA. The replicate fingerprints were initiated by using either 125 ng or 250 ng of RNA as template during first strand cDNA synthesis. Reaction conditions for first strand cDNA synthesis were 250 units of SuperScript II™ (GIBCO-BRL, Gaithersburg, MD) in 1X supplier's reaction buffer (25 mM Tris-HCl [pH=8.3], 37.5 mM KCl, 3.0 mM $MgCl_2$), 10 mM DTT, 400 μM each dNTP, and 2.0 μM permiscuous oligonucleotide in a 40 μl volume. The latter was incubated for 1 h at 37°C. Following first strand cDNA synthesis, the RNA was digested with RNase H and heat inactivated at 70°C as directed by the supplier.

One-tenth (4.0 μl) of the first strand cDNA reaction mixture was used in the fingerprinting PCR™ reaction. As many as ten different RNA fingerprints were generated from each first strand cDNA reaction. To the first strand cDNA, 36 μl of a PCR™ mix solution was added. The latter contained 50 mM Tris-Cl (pH=8.3), 50 mM KCl, 200 μM each dNTP, 1.0/μCi of α^{33} P-dCTP, 2.0 μM second permiscuous oligonucleotide and 1.0 unit of recombinant Taq DNA polymerase (GIBCO-BRL, Gaithersburg, MD). Note that the concentration of the first oligonucleotide is now slightly less that 200 nM. PCR™ fingerprinting was performed with one cycle of 94°C for 2:00 min, 48°C for 5:00 min then 72°C for 5:00 min. This was followed by 35 cycles of 94°C for 45 sec, 48°C for 1:30 min, and 72°C for 2:00 min. A final extension step of 72°C for 5:00 was performed. Next, 4.0 μl of the final PCR™ products were mixed with 6.0 μl of sequencing formamide dye solution and denatured by heating to 75°C for 5:00 min. Approximately 2.5 μl of the denatured PCR™ products in formamide dye was electrophoresed through a 6% polyacrylamide, 7M urea DNA sequencing gel. PCR™ products were visualized by autoradiography.

The two differentially appearing PCR™ amplified cDNA fragments identified in these studies that are the subjects of this report were termed UC331 and UC332. UC331 was

identified in a study in which the first permiscuous primer used in the reverse transcription reaction had the sequence 5′ ACGACTCACTATAAGCAGGA 3′ (SEQ ID NO:24). The second permiscuous primer that was used in the PCR™ fingerprinting reaction that identified UC331 was 5′ AACAGCTATGACCATCGTGG 3′ (SEQ ID NO:25). UC332 was identified in a study in which the first permiscuous primer used in the reverse transcription reaction had the sequence 5′ ACGACTCACTATGTGGAGAA 3′ (SEQ ID NO:26). The second permiscuous primer that was used in the PCR™ fingerprinting reaction that identified UC332 was 5′ AACAGCTATGACCCTGAGGA 3′ (SEQ ID NO:27). After autoradiography, bands that appeared differentially in fingerprinting reactions on the pooled total RNAs described above were cut out of the gels and reamplified by PCR™. The reamplified PCR™ products were directly sequenced using the Sequenase™ reagent system (Amersham Life Sciences, Inc., Arlington Heights, IL.).

The sequences of UC331 and UC332 were compared to those deposited in release 101 of GenBank (July 1997) using the Lasergene™ software package (DNAstar, Inc., Madison, WI). The DNA sequence of these cDNA fragments, when compared to the GenBank database, revealed that the mRNAs, from which these cDNA fragments were derived, were previously uncharacterized. Neither UC331 nor UC332 are genes whose products have been previously characterized as being significant in any physiological pathway, both UC331 and UC332 match sequences on the GenBank data base.

In the case of UC331, these matches are confined to ESTs. UC331 was identical within the limits of sequencing accuracy to several human EST sequences. The human EST sequences with high similarity to UC331 could be assembled into a virtual contig that predicts the sequence of a larger mRNA. The ends of the UC331 contig were then used to requery the EST data base whereby more ESTs were identified that extended the contig. This process was continued until the UC331 contig predicted a mRNA with an ORF and a poly-A tail. A description of the human ESTs that were used to construct the UC331 contig are provided in Table 9. The sequence of the UC331 contig and the ORF was identified at its 5′ end. A significant feature of this contig is that the ORF extends all the way to its 5′ end. This indicates that the UC331 mRNA extends further 5′ than is indicated by the contig constructed from the EST database.

TABLE 9
UC331 EST Distribution
Human

GB Accession Number	Tissue	Library
AA403120	Total Fetus	Soares
AA401845	Total Fetus	Soares
AA121473	Pregnant Uterus	Soares
AA121262	Pregnant Uterus	Soares
R22145[j]	Placenta	Soares
R22146[j]	Placenta	Soares
R30954[i]	Placenta	Soares
R31006[i]	Placenta	Soares
R32887[h]	Placenta	Soares
R31390[h]	Placenta	Soares
R67806[g]	Placenta	Soares
R67807[g]	Placenta	Soares
AA385620	Thyroid	TIGR
W37985	Parathyroid Tumor	Soares
W37986	Parathyroid Tumor	Soares
AA380401	Cell line (Supt)	TIGR
AA182471	Cell line (HeLa)	Stratagene (IMAGE)
AA181530	Cell line (HeLa)	Stratagene (IMAGE)
W31231	Senescent Fibroblasts	Soares
N22701	Normal Melanocyte	Soares
N31175	Normal Melanocyte	Soares
N34446	Normal Melanocyte	Soares
N34538	Normal Melanocyte	Soares
N36424	Normal Melanocyte	Soares
N36521	Normal Melanocyte	Soares
N42854	Normal Melanocyte	Soares
N44299	Normal Melanocyte	Soares

GB Accession Number	Tissue	Library
W56398	Normal Melanocyte	Soares
N66813	Normal Melanocyte	Soares
AA379996	Skin Tumor	TIGR
AA370040	Prostate Gland	TIGR
AA369851	Prostate Gland	TIGR
H08822[h]	Brain (Whole infant)	Soares
H08905[h]	Brain (Whole infant)	Soares
H19533	Brain (Whole Adult)	Soares
H21379[f]	Brain (Whole Adult)	Soares
H21421[f]	Brain (Whole Adult)	Soares
H24360[e]	Brain (Whole Adult)	Soares
H25176[e]	Brain (Whole Adult)	Soares
H38689	Brain (Whole Adult)	Soares
H38791	Brain (Whole Adult)	Soares
H39147[d]	Brain (Whole Adult)	Soares
H39148[d]	Brain (Whole Adult)	Soares
H45092[c]	Brain (Whole Adult)	Soares
H45054[c]	Brain (Whole Adult)	Soares
H49928	Brain (Whole Adult)	Soares
H50463	Brain (Whole Adult)	Soares
H51403[a]	Brain (Whole Adult)	Soares
H51444[a]	Brain (Whole Adult)	Soares
H52811[b]	Brain (Whole Adult)	Soares
H52774[b]	Brain (Whole Adult)	Soares
R85542	Brain (Whole Adult)	Soares
R84652	Brain (Whole Adult)	Soares
AA324855	Brain (Cerebellum)	TIGR
AA317211	Retina	TIGR
AA371911	Pituitary Gland	TIGR
AA302113	Endothelial Cells, Aorta	TIGR
AA247643	Fetal Heart	U. Toronto
W60049	Fetal Heart	Soares
W61359	Fetal Heart	Soares

GB Accession Number	Tissue	Library
AA243511	B-Cells	Soares
AA234769	Pooled; fetal heart, melanocytes, pregnant uterus	Soares
AA 158239	Pancreas	Stratagene (IMAGE)
AA 150565	Pancreas	Stratagene (IMAGE)
AA 160836	Pancreas	Stratagene (IMAGE)
H73822	Fetal Liver Spleen	Soares
N58180	Fetal Liver Spleen	Soares
W04414	Fetal Liver Spleen	Soares
N94254	Fetal Liver Spleen	Soares
N75996	Fetal Liver Spleen	Soares
N69644	Fetal Liver Spleen	Soares
T83329	Fetal Liver Spleen	Soares
T72755	Fetal Liver Spleen	Soares
T53976	Pooled Fetal Spleens	Soares
N76701	Multiple Sclerosis	Soares
N90814	Multiple Sclerosis	Soares
N63292	Multiple Sclerosis	Soares
N59233	Multiple Sclerosis	Soares
N53207	Multiple Sclerosis	Soares
N51545	Multiple Sclerosis	Soares
F22624	Skeletal Muscle.	CRIBI (Italy)

Note:Paired superscripts indicate opposite ends of the same cDNA clone.

When the human UC331 contig was used to query the GenBank database many mouse EST sequences were identified with significant similarity. This was especially true in the region spanning the putative ORF. The identified mouse ESTs were found to have areas of overlap and similarity with each other that permitted them to be assembled into a mouse UC331 virtual

contig in a process that was identical to that used to create the human contig. The mouse UC331 virtual contig was also observed to have an ORF at its 5′ end and a poly-A tail at its 3′ end. A description of the mouse ESTs that were used to construct this contig are provided in Table 10.

TABLE 10
Mouse

GB Accession Number	Tissue	Library	Clone #
AA027487	Placenta	Soares	459407 (5′)
AA023708	Placenta	Soares	456984 (5′)
AA023154	Placenta	Soares	456027 (5′)
AA024303	Placenta	Soares	458313 (5′)
W35948	Total Fetus	Soares	350258 (5′)
W 11581	Total Fetus	Soares	318665 (5′)
W36820	Total Fetus	Soares	336707 (5′)
AA002492	Mouse Embryo	Soares	426498 (5′)
AA097370	Mouse Embryo	Soares	493073 (5′)
AA014313	Mouse Embryo	Soares	468491 (5′)
AA450512	Beddington embryonic region	IMAGE	865186 (5′)
AA408179[L]	Embryo Ectoplacental Cone	Ko	C0025F09 (3′)
AA408261[L]	Embryo Ectoplacental Cone	Ko	C0025F09 (5′)
AA117174	T-cells	Stratagene	558134 (5′)
AA119346	Thymus	Soares	573567 (5′)
AA183195	Lymph Node	Soares	636222 (5′)
AA122933	Kidney	Barstead	579415 (5′)
AA423613	Mammary Gland	Soares	832219 (5′)

Note:Paired superscripts indicate opposite ends of the same cDNA clone.

When the MegAlign™ program of the Lasergene™ DNA analysis software package (DNAstar, Inc.) was used to compare the mouse and human UC331 contigs, the two contigs were predicted to represent mRNA species that were highly similar and nearly collinear throughout their lengths. This similarity was most striking in the region comprising the putative ORFs. Within the ORFs the mouse and human contigs, the DNA sequences are 89% identical. In the

predicted 3′ untranslated regions of the two contigs, the DNA sequence similarity falls to 73% with several small deletions and insertions. This higher degree of sequence similarity in the putative ORFs as compared to the proposed 3′ untranslated region is interpreted as evidence that the ORFs encode proteins on which natural selection constrains amino acid sequence divergence. Like the human UC331 contig, the mouse contig also encodes a putative ORF that extends all the way to its 5′ end. This provides additional support for the contention that the UC331 mRNA contains more sequences at its 5′ end than are represented by the EST based contigs presented here.

The ORFs of the mouse and human UC331 contigs were conceptually translated and the amino acid sequences were compared. The amino acid sequence of the human UC331 ORF was used to query the Swiss, PIR and Translation release 101 using the Lasergene™ software package. For the 157 amino acids for which this comparison is possible, the mouse and human sequences are collinear and identical at 151 positions (96%) with five of the six differences being conservative substitutions. This putative protein domain is highly acidic with 26 acidic and 17 basic amino acids. There were also 48 hydrophobic and 41 polar amino acids predicted. When either the predicted mouse or human UC331 amino acid sequences was compared to amino acid sequences in the public protein sequence data bases, no significant matches were found to any previously characterized vertebrate proteins. However, a significant match was observed to a putative protein, termed ZK353.1 (PIR Accession number S44654), encoded in the genome of the nematode, *Caenorhabditis elegans*. The mammalian amino acid sequence is similar and collinear with the C-terminal 157 amino acids of the putative *C. elegans* protein. Like the mammalian UC331 amino acid sequences, the C-terminal 157 amino acid sequence of the ZK353.1 is also highly acidic with 31 acidic and only 20 basic amino acids. Over the 203 amino acids for which a comparison can be made the ZK353.1 amino acid sequence is identical to the human or mouse sequence at 84 (41%) positions with many of the differences representing conservative substitutions.

The putative *C. elegans* protein, ZK353.1, has no currently known function. Its existence is predicted from the *C. elegans* genome sequencing effort (Sulston *et al.*, 1992). The polypeptide sequence for ZK353.1 is a conceptual translation of an area on the *C. elegans* chromosome III (GB accession number CELZK353). The predicted sequence for ZK353.1 is

548 amino acids long and includes an additional 371 amino acids that are N-terminal of the domain with similarity to the predicted amino acid sequence of UC331. If UC331 is the mammalian homolog of ZK353.1 and if UC331 is collinear with the *C. elegans* protein over its entire length, it could be expected that the ORF of UC331 would extend roughly an additional 1100 nucleotides 5′ of the sequence in SEQ ID NO:29. While it is likely that the UC331 ORF extends further 5′ than is accounted for in the virtual mouse and human UC331 contigs, Northern blot data from human poly-A plus RNA discussed below indicates that the human UC331 mRNA extends only about 350 nucleotides further 5′. This may indicate an error in interpreting the possible pattern of 'mRNA processing from the *C. elegans* sequence or indicate simply that the mammalian and nematode mRNAs and encoded proteins are significantly different from each other at their 5′ and N-terminal ends respectively.

To confirm that the human UC331 virtual contig accurately represented the sequence of an authentic mRNA, oligonucleotides were designed to direct the PCR™ amplification of large cDNA fragments predicted to be continuous from the virtual contig but which contain significantly more sequence than can be found in any single EST.

UC332 did not match any EST sequences but was identical to a portion of a previously sequenced full length cDNA with a GenBank accession number of D87451.

RELATIVE QUANTITATIVE RT-PCR™

Frequently, mRNAs identified by RNA fingerprinting or differential display as being differentially regulated turn out not to be so when examined by independent means. It is, therefore, critical that the differential expression of all mRNAs identified by RNA fingerprinting be confirmed as such by an independent methodology. To independently confirm the differential expression of UC331 in the peripheral blood of patients with recurrent metastatic cancer compared to the peripheral blood of healthy volunteers, two different formats for a relative quantitative RT-PCR™ were performed. The first format of this assay examined normalized pools of cDNA constructed by combining equal amounts of cDNA from various individuals representing similar physiologic states. In this study, a cDNA pool representing 8 healthy volunteers was compared to a pool representing 10 individuals with recurrent metastatic prostate cancer. A third pool representing 10 individuals with recurrent metastatic breast cancer was also

examined. The inclusion of the breast cancer patient samples in this study was made to determine if the mRNAs examined were being differentially regulated in the immune system in a response that was specific for prostate cancer or if the response was more general to metastatic cancer in general. Using these pools of cDNA as templates, triplicate PCR™ was performed. Each of the three replicates were terminated at a different cycle number of PCR™. This format of relative quantitative RT-PCR™ insures that the results taken for relative quantitation represent the PCRs™ when they are in the log linear portions of their amplification curves where such quantitation is most accurate.

Approximately 1.5-5.0 μg of DNA-free total RNA from the peripheral blood of healthy volunteers or patients with either metastatic prostate or breast cancer were converted into first strand cDNA using the *SuperScript™ Preamplification System* for First Strand cDNA Synthesis (GIBCO-BRL, Cat# 18089-011) following the directions provided by the supplier. These cDNAs were then normalized to contain equal concentrations of amplifiable cDNA by PCR™ amplification of β-actin cDNA using the primers 5′ GGAGCTGCCTGACGGCCAGGTCATC 3′ (SEQ ID NO:28) and 5′ GAAGCATTTGCGGTGGACGATGGAG 3′ (SEQ ID NO:9). A typical PCR™ program would be 94°C for 1:15 min, followed by 22 cycles of 94°C for 45 sec, 55°C for 45 sec and 72°C for 1:15 min. This was followed by final extension of 72°C for 5:00 min. PCR™ products were visualized by gel electrophoresis through 1.5% agarose TAE gels stained with ethidium bromide. Images of the gels were captured, digitized and analyzed using the IS-1000 Digital Imaging System (Alpha Innotech Corp.). The concentrations of the cDNAs were adjusted by adding various amounts of water to create cDNA stocks that contained equal concentrations of amplifiable β-actin cDNA. Typically, the cDNA derived from the reverse transcription of 5.0 μg of RNA resulted in enough normalized cDNA to perform 50-200 RT-PCR™ reactions.

Equal amounts of the normalized cDNA stock from individuals having the same disease state were pooled. Pools of cDNAs from healthy volunteers, patients with metastatic prostate cancer and metastatic breast cancer were produced. These pools were then examined by PCR™ for β-actin to determine that they contained equal amounts of amplifiable cDNA.

To demonstrate that all observations were made in the log-linear phase of the PCR™ amplification curve, a series of PCR™ reactions using different cycle number were performed on

each cDNA pool for each gene (primer pair) examined. Display of the PCR™ products on electrophoretic gels and analysis with the IS 1000 Digital Imaging System illustrates that the mass of the PCR™ products is increased exponentially with increasing cycle number, confirming that the observed results are in the log-linear portion of the PCR™ amplification curve.

Relative quantitative RT-PCR™ showing near equal amounts of amplifiable β-actin cDNA in three pools cDNA. Pools of normalized cDNAs were constructed from peripheral blood RNAs from eight healthy volunteers, ten individuals with recurrent metastatic prostate cancer, or ten individuals with recurrent metastatic breast cancer. Three separate PCR™ reactions were performed on each pool of cDNA. PCR™ was terminated at differing cycle numbers (cycle 22, cycle 24, and cycle 26), and the products were visualized by electrophoreses and ethidiumn bromide staining. Images were captured and quantitated using a digital image analysis system. At all three cycle numbers examined, there are relatively similar band intensities representing the three cDNA pools and increasing band intensity with increasing cycle number, verifying that the observations are being made in the log linear range of the amplification curves. Similar band intensities indicate similar relative concentrations of β-actin mRNA in the RNAs from individuals from which these cDNA pools were constructed.

The oligonucleotides used in the relative quantitative RT-PCR™ studies that independently confirmed the differential expression of UC331 were designed from the sequence in the human UC331 virtual contig. These UC331 specific oligonucleotides had the sequences of 5′ CTGGCCTACGGAAGATACGACAC 3′ (SEQ ID NO:31) and 5′ ACAATCCGGAGGCATCAGAAACT 3′ (SEQ ID NO:32). These oligonucleotides direct the amplification of a 277 nucleotide long PCR™ product that is specific for UC331. The oligonucleotides used in the relative quantitative RT-PCR™ studies that independently confirmed the differential expression of UC332 were designed using the sequences of the cDNA with the GenBank accession number D87451. These UC332 specific oligonucleotides had the sequences 5′ AGCCCCGGCCTCCTCGTCCTC 3′ (SEQ ID NO:33) and 5′ GGCGGCGGCAGCGGTTCTC 3′ (SEQ ID NO:34). These oligonucleotides direct the amplification of a 140 nucleotide long PCR™ product that is specific for UC332.

The results for relative levels of β-actin expression contrasts sharply with those observed when oligonucleotide primers specific for UC331 were used to direct PCR™ amplification (FIG.

7). At 25 cycles of PCR™, clear bands are visible in the lanes representing the pools of cDNA from peripheral blood of patients with either metastatic breast or prostate cancer. In the lane representing the peripheral blood of healthy volunteers, only a very faint band is present. At 28 cycles of PCR™, the band intensities representing all three pools are brighter than they were at 25 cycles, but the relative increase in intensity of the bands representing the metastatic cancer patient pools compared to the healthy volunteers remains the same as was observed at 25 cycles of PCR™. This indicates that these observations are being made in the log linear range of the PCR™ amplification curves. At 31 cycles of PCR™, there is still an increase in the intensity of the bands representing the pools of metastatic cancer patients compared to the pool representing the healthy volunteers, but a quantitative analysis of these bands indicates that the PCRs™ have left the log linear range of their amplification curves. Quantitation of the data for 25 and 28 cycles of PCR™ independently confirms that UC331 mRNA is differentially regulated and is roughly seven fold more abundant in the peripheral blood leukocytes of the average patient with either recurrent metastatic prostate cancer or breast cancer than in the peripheral blood leukocytes of healthy volunteers.

The second format of relative quantitative RT-PCR™ used to examine the differential expression of UC331 examined the relative abundance of UC331 mRNA in the peripheral blood of healthy individuals or individuals with recurrent metastatic cancer. The individuals examined in this study were the same as those whose cDNAs were combined to construct the pools examined as described above. Using the information obtained from the pooled cDNA study to predict at what PCR™ cycle numbers relative quantitative RT-PCR™ would be most informative, these individuals were examined for the relative abundance of β-actin and UC331 mRNAs present in their peripheral blood leukocytes. PCR™ was for 22 cycles. All individuals examined contain roughly equal amounts of amplifiable β-actin cDNA. Some of the differences in β-actin band intensity observed in this study are probably due to the internal variation inherent of this study. Results from studies designed to quantitate this internal variation indicate that identical replicates of a β-actin PCR™ can be expected to vary in the intensity of product bands with a standard deviation of ±15%.

Relative quantitative RT-PCR™ of UC331 cDNA was conducted using reverse transcribed from RNA isolated from the peripheral blood of eight healthy volunteers (group N),

ten individuals with recurrent methstatic prostate cancer (group P), or ten individuals with recurrent metastatic breast cancer (group B). PCR™ was for 30 cycles. As was seen in the study using the pooled cDNAs, the results of the relative quantitative RT-PCR™ for UC331 using cDNA from individuals contrasts sharply with that observed for β-actin. The intensity of the band representing the abundance of the UC331 mRNA in peripheral blood leukocytes was greater for all of the patients with either metastatic prostate or breast cancer as compared to the intensity of the UC331 band representing the mRNA level in the peripheral blood leukocytes of healthy volunteers. Therefore, the elevated UC331 mRNA levels indicated by the relative quantitative RT-PCR™ results using the pooled cDNA templates was caused by an elevated mRNA level in all individuals comprising the pools and not from a subset of individuals with very high elevations in UC331 mRNA levels. This study is a second independent confirmation of the differential expression of the UC331 mRNA.

As is indicated by the wide distribution of tissues from which the ESTs used to assemble the UC331 contigs (Table 9), UC331 is widely expressed in many tissue and cell types. However, because most of ESTs comprising UC331 are from normalized libraries, little information can be gained from this data on the relative abundance of the UC331 mRNA in different tissues. Also, while the extension of the ORFs of the mouse and human UC331 contigs all the way to their 5′ ends and the similarity of mammalian UC331 mRNAs to a much larger putative *C. elegans* mRNA both predict that the mammalian UC331 mRNA extends even further 5′, the exact size of the UC331 mRNA was unknown. To address all of these issues, a Northern blot of poly-A plus RNA from eight different human tissues was probed with the 850 nucleotide long RT-PCR™ product described above labeled with ^{32}P. Approximately 2.0 μg of poly-A plus RNA from spleen, thymus, prostate, testis, ovary, small intestine, colon, and peripheral blood leukocytes were loaded in each lane. UC331 mRNA is expressed in all eight human tissue and cell types. Size standards indicate a message size of approximately 1.75 kb. Interestingly, UC331 is least abundant in peripheral blood leukocytes but is highly expressed in the thymus, demonstrating a difference in expression between cells of different developmental stages in the immune system. UC331 is most abundantly expressed in the testes. The UC331 mRNA is about 1.75 kb which indicated that the mRNA only extends about 350 nucleotides further 5′ than is accounted for by the virtual contig shown in SEQ ID NO:29. The translation product of the

virtual contig is shown in SEQ ID NO:30. Clearly, the putative *C. elegans* mRNA extends much more 5′ than do the mammalian mRNA species.

The other gene identified as being differentially regulated in this RNA fingerprinting study was UC332. UC332 was analyzed in much the same way as UC331 was. When the sequence of the cDNA fragment from the RNA fingerprinting gel representing UC332 was used to query GenBank, no ESTs were identified. The sequence of the UC332 cDNA fragment did, however, identify a sequence of a full length cDNA, KA000262 (GB:accession number D87451). The sequence of KA000262, (hereafter referred to interchangeably with the name, UC332) was determined as part of a project to examine previously unidentified mRNAs expressed in the bone marrow myeloblast cell line, KG-1 (Nagase *et al.*, 1996). This mRNA contains an ORF encoding a putative protein with 761 amino acid sequence. Perhaps the most striking feature of this polypeptide sequence is the appearance of a C3HC4 RING zinc finger or RING finger motif (Freemont, 1993) located between amino acids 175 and 216. The RING finger domain binds two zinc ions in a conserved structure that has been resolved (Barlow *et al.*, 1994). RING finger domains have been identified in dozens of proteins derived from eukaryotes as diverse as yeasts, flies, birds, nematodes and humans. In most of these cases, the RING finger containing proteins have been shown to be essential for some important biological process although the these processes vary considerably one from another. Among these mammalian encoded RING finger proteins are several genes implicated in the ontogeny of cancer including the *ret* viral oncogene (Takahashi *et al.*, 1988) and bmi-1, a gene whose product collaborates with myc induced transformation (Haupt *et al.*, 1991). The BRCA-1 tumor suppressor gene involved in hereditary breast and ovarian cancer susceptibility contains a RING finger domain (Miki *et al.*, 1994), and MAT-1, a novel 36 kDa RING finger protein, is required for the assembly of enzymatically active CDK7- cyclin H complexes (Tassan *et al.*, 1995). A comparison of the RING finger domains of UC332 and various representative members of this group, including BRAC1, rpt-1, Traf5, HT2A, MAT1, rfp, bmi-1, CRZF, and neu, indicates the RING finger domain of UC332 is slightly more similar to those found in the tumor suppressor gene, BRCA1, and the T cell repressor of transcription protein, rpt-1. However, BRCA1 and rpt-1 are more similar to each other than they are to UC332.

Proteins with RING finger motifs exhibit heterogeneity in their subcelluar localizations. Some, that are important regulators of differential gene regulation, localize to the cell nucleus. When the amino acid sequence of UC332 was scanned for evidence of subcellular localization, two domains were identified that contained sequences for putative nuclear localization signals (NLS). NLS are highly basic stretches of six are more amino acids of which at least four are basic that tend to be flanked by acidic amino acids and/or prolines (Boulikas, 1994). Both of the putative NLS in UC332 longer and more basic than the minimum requirements for the consensus NLS motif. The first of these putative NLS motifs occurs between amino acid 548 and 567. Within this domain, 13 of 19 amino acids are basic. In fact, this domain could be viewed as two NLS in tandem separated by two glutamic acid residues. If divided this way, the first NLS domain would have 8 of eleven positions as basic amino acids while the second motif would have 5 of 6 amino acids being basic. The second NLS motif in UC332 is located near the C-terminal end between positions 739 and 750 in the amino acid sequence. This domain has 8 of 12 amino acids as basic residues with a core of 5 consecutive lysines and arginines. The presence of these putative NLS in the amino acid sequence of UC332 suggest the possibility that UC332 plays an important role in regulating the expression of other genes. Finally, the amino acid sequence of UC332 lacks a signal sequence for cellular export or an obvious hydrophobic transmembrane domains.

To independently verify that UC332 mRNA is more abundant in the peripheral blood leukocytes of patients with recurrent metastatic cancer as compared to the peripheral blood leukocytes of healthy volunteers, relative quantitative RT-PCR™ was performed using the same cDNAs and formats as were used to investigate the differential regulation of UC331. A relative quantitative RT-PCR™ study using UC332 specific oligonucleotide primers and cDNA pools as templates was conducted. At 25 and 28 cycles of PCR™, the amplified DNA band representing the relative abundance of the UC332 mRNA is stained more intensely for those reactions that used cDNA template pools constructed from the peripheral blood leukocyte RNA isolated from metastatic prostate and breast cancer patients as compared to a similar pool constructed from RNA from healthy volunteers. Quantitation of this image using the IS-1000 Digital Imaging System (Alpha Innotech, Inc.) indicates that UC332 mRNA is roughly 5 times more abundant in

the peripheral blood leukocytes of metastatic cancer patients compared to healthy volunteers. At 31 cycles of PCR™, the reactions have left the log linear range of their amplification curves.

In a second relative quantitative RT-PCR™ study using UC332 specific oligonucleotide primers, peripheral blood leukocyte cDNA from the individuals that comprised the pools from the peripheral blood of eight healthy volunteers, ten individuals with recurrent metastatic prostate cancer, or ten individuals with recurrent metastatic breast cancer were examined separately. PCR™ was for 26 cycles. The results of this study are similar to those obtained when the pooled cDNAs were used as PCR™ templates. All of the cancer patients had higher levels of UC332 mRNA in their peripheral blood leukocytes than did any of the healthy volunteers.

In this study, the inventors showed that UC332, encoding a RING finger protein, is up regulated in the peripheral blood leukocytes of patients with either recurrent metastatic breast or prostate cancer. From the literature, RING finger proteins have been shown to participate in the regulation of several important lymphocytic processes (Patarca *et al.*, 1988; Fridell *et al.*, 1995; Takeuchi *et al.*, 1996; van Arsdale *et al.*, 1997; Nakano *et al.*, 1996). The observed differential regulation of the RING protein encoding mRNA, UC332, in the immune response of patients with metastatic breast or prostate cancer strongly suggests that UC332 participates in regulating this immune response.

All of the compositions and methods disclosed and claimed herein may be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this disclosure have been described in terms of preferred embodiments, it is apparent that variations may be applied to the composition, methods and in the steps or in the sequence of steps of the method described herein without departing from the concept, spirit and scope of the invention.

More specifically, it is apparent that certain agents which are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the disclosure as defined by the appended claims.

UC325-1 is derived from the IL-8 gene (Genebank Accession #M28130). UC325-1 and UC325-2, an alternatively spliced form that includes the third intron of the IL-8 primary transcript,

are transcribed from the IL-8 gene. Our definition of IL-8 gene products means all mRNAs transcribed from the IL-8 gene, the polypeptides encoded by those mRNAs and their post-translationally processed protein products.

Those practiced in the art will realize that there exists naturally occurring genetic variation between individuals. As a result, some individuals may synthesize IL-8 gene products that differ from those described by the sequences entailed in the Genebank number listed above. We include in our definition of IL-8, those products encoded by IL-8 genes that vary in sequence from those described above. Those practiced in the art will realize that modest variations in DNA sequence will not significantly obscure the identity of a gene product as being derived from the IL-8 gene.

REFERENCES

The following references, to the extent that they provide exemplary procedural or other details supplementary to those set forth herein, are specifically incorporated herein by reference.

Abbondanzo *et al.*, *Breast Cancer Res. Treat.*, 16: 182(#151), 1990.

Adams *et al.*, 1995.

Adams, Kelley, Gocayne, Dubnick, Polymeropoulos, Xiao, Merril, Wu, Olde, Moreno, Kerlavage, McCombie, Venter, "Complementary DNA sequencing: Expressed sequence tags and he human genome," *Science*, 252:1651-1656, 1991.

Adams, Kerlavage, Fleischmann, Fuldner, *et al.*, "Initial assessment of the human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence," *Nature 377: Supplement: The Genome Directory*, pages 3-174.

Alcaraz *et al.*, *Cancer Res.*, 55:3998-4002, 1994.

Allred *et al.*, *Breast Cancer Res. Treat.*, 16:182(#149), 1990.

An *et al.*, *Proc. Amer. Assn. Canc. Res.*, 36:82, 1995.

Andrawis *et al.*, *Proc. Am. Urol. Assn.* 155(5): 512A, 1996.

Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor Press, Cold Spring Harbor, New York, 1988.

Barinaga, *Science*, 271: 1233, 1996.

Barlow, Luisi, Milner, Elliott, Everett, "Structure of the C3HC4 Domain by H-nuclear Magnetic Resonance Spectroscopy," *J. Mol. Biol.*, 237:201-211, 1994.

Barry *et al.*, *Urology*, 46:2-14, 1995.

Beck, Kattan and Miles, *J. Urol.*, 152:1894-99, 1994.

Bellus, *J. Macromol. Sci. Pure Appl. Chem*, A31(1): 1355-1376, 1994.

Bittner *et al.*, *Methods in Enzymol*, 153: 516-544, 1987.

Frohman, *PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS*, Academic Press, N.Y., 1990.

Fujichima *et al.*, *Cytometry*, 24:382-389, 1986.

Garnick, *Scientific Amer.* 270: 72-81, 1994.

Gefter *et al.*, *Somatic Cell Genet.*, 3: 231-236, 1977.

Gingeras *et al.*, PCT Application WO 88/10315.

Goding, *In: Monoclonal Antibodies: Principles and Practice*, 2d ed., Orlando, Fla., Academic Press, pp. 60-61, 65-66, 71-74, 1986.

Haupt, Alexander, Barri, Klinken, Adams, "Novel zinc finger gene implicated as *myc* collaborator by retrovirally accelerated lymphomagenesis in *Ep.-myc* transgenic mice," *Cell*, 65:753-763, 1991.

Hedges *et al.*, *Cytokine* 8:686-697, 1996.

Hess *et al.*, *J. Adv. Enzyme Reg.*, 7:149, 1968.

Heston, personal communication, CapCURE meeting, Lake Tahoe, CA, 1996.

Hitzeman *et al.*, *J. Biol. Chem.*, 255:2073, 1980.

Holland *et al.*, *Biochemistry*, 17:4900, 1978.

Holmes *et al.*, *Prostate*, 27:25-29, 1996.

Horoszewicz, Kawinski and Murphy, *Anticancer Res.*, 7:927-936, 1987.

Hosmer and Lemeshow, *In: Applied Logistic Regression*, New York, John Wiley & Sons, 1989.

Huang *et al.*, *Prostate*, 23: 201-212, 1993.

Ikei *et al.*, *Cytokine* 4: 581-584, 1992.

IMAGE ref.

Innis *et al.*, *PCR Protocols*, Academic Press, Inc., San Diego CA, 1990.

Inouye *et al.*, *Nucleic Acids Res.*, 13:3101-3109, 1985.

Isaacs *et al.*, *Cancer Res.*, 51:4716-4720, 1991.

Isaacs *et al.*, *Seminars in Oncology*, 21:1-18, 1994.

Israeli *et al.*, *Cancer Research*, 54:1807-1811, 1994.

Issacs Reference

Johnson *et al.*, *In: BIOTECHNOLOGY AND PHARMACY*, Pezzuto *et al.*, Eds., Chapman and Hall, New York, 1993.

Jones, *Genetics*, 85:12, 1977.

Kingsman *et al.*, Gene, 7: 141, 1979.

Koch *et al.*, *Science*, 258:1798-1801, 1992.

Kohler and Milstein, *Eur. J. Immunol.*, 6: 511-519, 1976.

Kohler and Milstein, *Nature*, 256:495-497, 1975.

Kurdowska *et al.*, *J. Immunol.*, 154:1930-1940, 1997.

Kwoh *et al.*, *Proc. Nat. Acad. Sci. USA*, 86: 1173, 1989.

Liang and Pardee, "Differential display of eukaryotic messenger RNA by means of polymerase chain reaction," *Science*, 257:967-971.1992.

Liang and Pardee, *Science*, 257: 967-971, 1992.

Liang and Pardee, U.S. Patent 5,262,311, 1993.

Liang *et al.*, *Cancer Res.*, 52:6966-6968, 1992.

Lifton, *Science*, 272: 676, 1996.

Lilja *et al.*, *Clin. Chem.*, 37:1618-1625, 1991.

Lopez-Nieto and Nigam, "Selective amplification of protein-coding regions of large sets of genes using statistically designed primer sets," *Nature Biotechnology*, 14:857-861, 1996.

Lowy *et al.*, *Cell*, 22:817, 1980.

Macoska *et al.*, *Cancer Res.*, 54:3824-3830, 1994.

Marley *et al.*, *Urology*, 48(6A):16-22, 1996.

McClelland *et al.*, *Nucl. Acids Res.* 22: 4419-4431, 1994.

McCormack *et al.*, *Urology*, 45:729-744, 1995.

Miki *et al.*, *Science*, 266: 66-71, 1994.

Miki, Swensen, Shatmck-Eiden, Futreal, Harshman, Tavigian, Liu, Cochran, Bennett, Ding, Bell, Rosenthai, Hussey, Tran, McClure, Frye, Hattier, Phelps, Haugen-Strano, Katcher, Yakumo, Gholami, Shaffer, Stone, Bayer, Wray, Bogden, Dayananth, Ward, Tonin, Narod, Bristow, Norris, Helvering, Morrison, Rosteck, Lie, Barrett, Lewis, Neuhauses, Cannon-Albright, Goldgar, Wiserman, Kamb, Skolnick, "A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1," *Science*, 266: 66-71, 1994.

Miller *et al.*, PCT Application WO 89/06700.

Mok *et al.*, *Gynecol. Oncol.*, 52: 247-252, 1994.

Montie, *Urology*, 44: 2-8, 1994.

Morahan *et al.*, *Science* 272: 1811, 1996.

Morise *et al.*, *Med. Dec. Making*, 16(2):133-142, 1996.

Morton *et al.*, *Cancer Res.*, 53:3585-3590, 1993.

Morton *et al.*, *In: CANCER MEDICINE* (3rd Ed.), Holland, J.F., Frei III, E., Bast Jr., C.C. (eds), Lea and Febiger, Philadelphia, PA, pp. 1793-1824, 1993.

Mulligan *et al.*, *Proc. Nat'l Acad. Sci. USA*, 78:2072, 1981.

Murphy *et al.*, *Cancer*, 78: 809-818, 1996.

Murphy *et al.*, *Prostate*, 26:164-168, 1995.

Nagase, Seki, Ishikawa, Ohira, Kawarabayasi, Ohara, Tanaka, Kotani, Miyajima, Nomura, "Prediction of the coding sequences of unidentified human genes. VI. The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by analysis of cDNA clones from cell line KG-1 and brain," *DNA Res.*, 3(5):321-329, 1996.

Nakamura *et al.*, *In: Handbook of Experimental Immunology* (4th Ed.), Weir, E., Herzenberg, L.A., Blackwell, C., Herzenberg, L. (eds), Vol. 1, Chapter 27, Blackwell Scientific Publ., Oxford, 1987.

Nakano, Oshima, Chung, Williams-Abbott, Wares, Yagita, Okumura, "TRAF5, an activator of NF-κB an putative signal transducer for lymphotoxin-β receptor," *J. Biol. Chem.*, 271:14661-14664, 1996.

O'Hare *et al.*, *Proc. Nat'l Acad. Sci. USA*, 78:1527, 1981.

Oesterling *et al.*, *J. Urol.*, 154:1090-1095, 1995.

Oesterling, *Contemporary Urology*, 8:76-92, 1996.

Ohara *et al.*, *Proc. Nat'l Acad. Sci. USA*, 86:5673-5677, 1989.

Papsidero *et al.*, *Cancer Research*, 40:2428-2432, 1980.

Parker *et al.*, *CA Cancer J. Clin.*, 65:5-27, 1996.

Parker, Tong, Bolden, Wingo, "Cancer statistics," *CA Cancer J. Clin.*, 65:5-27, 1996.

Partin and Oesterling, (Eds.) *Urology*, 48(6A) supplement: 1-87, 1996.

Partin and Oesterling, *J. Urol.*, 152:1358-1368, 1994.

Partin *et al.*, *Cancer Res.*, 53:744-746, 1993.

Patarca, Scwartz, Singh, Kong, Murphy, Anderson, Sheng, Singh, Johnson, Guarnagia, Durfee, Blattner, Cantor, "rpt-1, an intracellular protein from helper/inducer T cells that regulates gene expression of interleukin 2 receptor and human immunodeficiency virus type 1," *Proc. Natl. Acad. Sci. USA*, 85:2733-2737, 1988.

Pearsons *et al.*, *J. Urol.*, 150:120-125, 1993.

Pettersson *et al.*, *Clin. Chem.*, 41:1480-88, 1995.

Piironen *et al.*, *Clin. Chem.* 42: 1034-1041, 1996.

Price, Chang, Smith, Bockheim, Laughon, "The Drosophila neutralized gene encodes a C3HC4 zinc finger," *EMBO J.*, 12:2411-2418, 1993.

Qiao *et al.*, *Biochem. Biophys. Res. Comm.*, 201:581-588, 1994.

Ralph *et al.* 1993.

Ribas de Pouplana and Fothergill-Gilmore, *Biochemistry*, 33:7047-7055, 1994.

Rieber, M. and Rieber, M. S., *Cell Growth Diff.*, 5:1339-1346, 1994.

Sager *et al.*, *FASEB J.*, 7:964-970, 1993.

Saito *et al.*, *Biochem. Biophys. Res. Commun.*, 200:378, 1994.

Sambrook *et al.*, (ed.), *MOLECULAR CLONING*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

Schadendorf *et al.*, *J. Immunol.*, 151:2667-2675, 1993.

Scheibenbogen *et al.*, *Melanoma Research*, 5:179-181, 1995.

Schroder, J.M., Sticherling, M., Henneicke, H. H., Preissner, W. C., Christophers, E. *J. Immunol.* 144:2223-2232, 1990.

Scott *et al.*, *Molec. Cell. Biol.*, 13:2247-2257, 1993.

Sidransky *et al.*, *Science*, 252: 706-709, 1991.

Silver *et al.*, *Clin. Cancer Res.*, 3: 81-85, 1997.

Slamon *et al.*, *Science*, 224:256-262, 1984.

Slamon *et al.*, *Science*, 244: 707-712, 1989.

Smith, U.S. Patent No. 4,215,051

Stamey, *Urology*, 45(2):173-184, 1995.

Steinberg and Colla, *CART: A supplementary module for SYSTAT*, Evanston, IL, SYSTAT Inc., 1992.

Stenman *et al.*, *Cancer Research*, 51:222-226, 1991.

Stinchcomb *et al.*, *Nature*, 282:39, 1979.

Stone *et al.*, *Urology*, 44: 18-25, 1994.

Sulston, Du, Thomas, Wilson, Hillier, Staden, Halloran, Green,, Thierry-Mieg, Qiu, Dear, Coulson, Craxton, Durbin, Berks, Metzstein, Hawkins, Ainscough, Waterston, "The *C. elegans* genome sequencing project: A beginning," *Nature*, 356:37-41, 1992.

Sun and Cohen, *Gene*, 137:127-132, 1993.

Szybalska *et al.*, *Proc. Nat'l Acad. Sci. USA*, 48: 2026, 1962.

Takahashi *et al.*, *Cancer Res.*, 54:3574-3579, 1994.

Takahashi, Inaguma, Hiai, Hirose, "Developmentally regulated expression of human "Finger"-containing gene encoded by the 5' half of the *ret* transforming gene," *Mol. Cell. Biol.*, 8:1853-1856, 1988.

Takeuchi, Roth, Goeddel, "Anatomy of TRAF2: Distinct domains for Nuclear Factor-κB activation and association with tumor necrosis factor signaling proteins," *J. Biol. Chem.*, 271:19935-19942, 1996.

Tan, Waterham, Veenhuis, Gregg, "The *Hansenula polymorpha* PER8 gene encodes a novel peroxisomal integral membrane protein involved in proliferation," *J. Cell Biol.*, 128:307-319, 1995.

Taparowsky *et al.*, *Nature*, 300: 762-764, 1982.

Tassan, Jaquenoud, Fry, Frutiger, Hughes, Nigg, "*In vitro* assembly of a functional human CDK7-cyclin H complex requires MAT1, a novel 36 kDa RING finger protein," *EMBO J.*, 14:4508-5617, 1995.

Thanos and Maniatis, "NF-κB: A lesson in family values," *Cell*, 80:529-532," 1995.

TIGR reference

Tranque, Crossin, Cirelli, Edelman, Mauro, "Identification and characterization of a RING zinc finger gene (C-RZF) expressed in chicken embryo cells," *Proc. Natl. Acad. Sci. USA*, 93:3105-3109, 1996.

Tschemper *et al.*, *Gene*, 10:157, 1980.

Tummuru, Sharma and Blaser, *Molec. Microbiol.* 18: 8670876, 1995.

Umbas *et al.*, *Cancer Res.*, 52:5104-5109, 1992.

VanArsdale, VanArsdale, Force, Walter, Mosialos, Kieff, Reed, Ware, "Lymphotoxic-β receptor signaling complex: Role of tumor necrosis factor receptor-associated factor 3 recruitment in cell death and activation of nuclear factor κB," *Proc. Natl. Acad. Sci. USA*, 94:2460-2465, 1997.

Vinante *et al.*, *Leukemia*, 7:1552-1556, 1993.

Visakorpi *et al.*, *Am. J. Pathol.*, 145:1-7, 1994.

Vishwanatha *et al.*, *Carcinogenesis*, 14:2575-2579, 1993.

Vollmer, *Am. J. Clin. Pathol.*, 105:115-126, 1996.

Walker *et al.*, *Proc. Nat'l Acad. Sci. USA*, 89:392-396, 1992.

Walsh and Worthington, *The Prostate*, Chap. 3, Johns Hopkins University Press, 1995.

Wan, Sharp, Poirier, Wagaman, Chambers, Pyati, Hom, Galindo, Huvar, Peterson, Jackson, Erlander, "Cloning differentially expressed mRNAs," *Nature Biotechnology*, 14:1685-1691, 1996.

Watson *et al.*, *Cancer Res.*, 54:4598-4602, 1994.

Webb and Lin, *Invest. Urol.*, 17:401-404, 1980.

Welsh and McClelland, *Nucl. Acids Res.*, 18:7213-7218, 1990.

Welsh *et al.*, *Nucleic Acids Res.*, 20:4965-4970, 1992.

Welsh, Chada, Dalal, Chang, Ralph, McClelland, "Arbitrarily primed PCR™ fingerprinting of RNA," *Nucl. Acids Res.*, 20: 4965-70, 1992.

Whittemore, *Biometrika*, 82(1):57-67, 1995.

Wigler *et al.*, *Cell*, 11: 223, 1977.

Wigler *et al.*, *Proc. Nat'l Acad. Sci. USA*, 77:3567, 1980.

Wong *et al.*, *Int. J. Oncol.*, 3:13-17, 1993.

Wu *et al.*, *Genomics*, 4:560, 1989.

Yoon, Lee, Jongeward, Steinberg, "Similarity of sli-1, a regulator of vulval development in *C. elegans*, to the mammalian proto-oncogene c-cbl," *Science*, 269:1102-1105, 1995.

Yoshimura, T., Robinson, E. A., Appella, E., Matsushima, K., Showalter, S.D., Skeel, A., and Leonard, E.J. *Molecular Immunology.* 26: 87-93, 1989.

Zhang *et al.*, *Clin. Chem.*, 41(11):1567-73, 1995.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: UROCOR, Inc.
(B) STREET: 800 Research Parkway
(C) CITY: Oklahoma City
(D) STATE: Oklahoma
(E) COUNTRY: USA
(F) POSTAL CODE (ZIP): 73104

(ii) TITLE OF INVENTION: DIAGNOSIS OF DISEASE STATE USING mRNA PROFILES

(iii) NUMBER OF SEQUENCES: 34

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 253 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GGCAGGGGCT TGTGACTCTA AGATGGCTTC ATTCACATGC CTAGGGCCTC AGTAGGATGA 60

CTGGCATGGC CCTGGAAAAC TGCGAAGTCT TCTCTCTGTG CAAACTTTCA CCTGGACTTT 120

TTATATGATT CTGGAAGTAT TCCAAGAAGG CAAAAGTAAA AACTGCAAAG CGTCTTAAAA 180

TAGAAGTTCA GAAGCCACAT TATATCACTT CTGTTGCATT CTATCAAAGC AAGTCACAAG 240

CCCCTGCCAA TCA 253

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 183 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

CACACACTCC CCCATTCTGA GCCCCAAGAG GCTCATCCCT AAGGATGTCC AGAGATCCAA 60

GTGCAGAAGG AGAATGTGGT GAGGCTATTT ATTCCCCCAG TGCCTTCCCT GCTGGGCTAT 120

GGATGAACAG TGGCTGACTT CATCTAGGAA AGAGCTATGG CTTCTGTCTC CTGGAGCTCA 180

CCA 183

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 387 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GGTGAGCCCC AGGAGACAGA AGAGATATGA GGAAATTGTT AAGGAAGTCA GCACTTACAT 60

TAAGAAAATT GGCTACAACC CCGACACAGT AGCATTTGTG CCAATTTCTG GTTGGAATGG 120

TGACAACATG CTGGAGCCAA GTGCTAACAT GCCTTGGTTC AAGGGATGGA AAGTCACCCG 180

TAAGGATGGC AATGCCAGTG GAACCACGCT GCTTGAGGCT CTGGACTGCA TCCTACCACC 240

AACTCGTCCA ACTGACAAGC CCTTGCGCCT GCCTCTCCAA GGATGTTCTT ACAAAATTGG 300

TGGTATTGGT ACTGTTCCCT GTTTGGCCGA ATTGGAAAAC TGGTGTTCCT CCAAACCCCG 360

GTTATGGTGG GTTTCCTCCT CCTTGGA 387

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 366 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GGGCGGAACA AGGGAGCGCT AAAAGGAAAT TAGGATGTCA GGTGCATAAA GGAACATAAT 60

TCCAAAACCT TTCCAAACCC CAAATTTATT CAAAGGAACT GAGGAGTGGA TTGAGGAGTG 120

GACCAACACT GGCGCCAAAC ACAGAAATTA TTGTAAAGCT TTCTGATGGA AGAGAGCTCT 180

GTCTGGGCCC CAAGGAAAAC TGGGTGCAGA GGGTTGTGGA GAAGTTTTTG AAGAGGGCTG 240

AGAATTCATA AAAAAATTCA TTCTCTGTGG TATCCAAGAA TCAGTGAAGA TGCCAGTGAA 300

ACTTCAAGCA AATCTACTTC AACACTTCAT GTATTGTGTG GGTCTGTTGT AGGGTTGCCA 360

GTTGTT 366

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 598 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GCTTGGGCCC CAAGGAAAAC TGGGTGCAGA GGGTTGTGGA GAAGTTTTTG AAGAGGTAAG 60

TTATATATTT TTGAATTTAA AATTTGTCAT TTATCCGTGA GACATATAAT CCAAAGTCAG 120

CCTATAAATT TCTTTCTGTT GCTAAAAATC GTCATTAGGT ATCTGCCTTT TTGGTTAAAA 180

AAAAAAGGAA TAGCATCAAT AGTGAGTGTG TTGTACTCAT GACCAGAAAG ACCATACATA 240

GTTTGCCCAG GAAATTCTGG GTTTAAGCTT GTGTCCTATA CTCTTAGTAA AGTTCTTTGT 300

CACTCCCAGT AGTGTCCTAT GTTAGATGAT AATGTCTTTG ATCTCCCTAT TTATAGTTGA 360

GAATATAGAG CATGTCTAAC ACATGAATGT CAAAGACTAT ATTGACTTTT CAAGAACCCT 420

ACTTTCCTTC TTATTAAACA TAGCTCATCT TTATATTGTG AATTTTATTT TAGGGCTGAG 480

AATTCATAAA AAAATTCATT CTCTGTGGTA TCCAAGAATC AGTGAAGATG CCAGTGAAAC 540

TTCAAGCAAA TCTACTTCAA CACTTCATGT ATTGTGTGGG TCTGTTGTAG GGTTGCCA 598

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CGCCTCAGGC TGGGGCAGCA TT 22

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

ACAGTGGAAG AGTCTCATTC GAGAT 25

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

CGAGCTGCCT GACGGCCAGG TCATC 25

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GAAGCATTTG CGGTGGACGA TGGAG 25

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

TGCAAACTTT CACCTGGACT T 21

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

CTTGTGACTT GCTTTGATAG AATG 24

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GACAACATGC TGGAGCCAAG TGC 23

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

ACCACCAATT TTGTAAGAAC ATCCT 25

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

TGTCCAGAGA TCCAAGTGCA GAAGG 25

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

GAGCTCCAGG AGACAGAAGC CATAG 25

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

GGGCCCCAAG GAAAACT 17

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

TGGCAACCCT ACAACAGAC 19

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

GGGCCCCAAG GAAAACT 17

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

TGGCAACCCT ACAACAGACC 20

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

ACATTGAAGC ACTCCGCGAC 20

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

AGAGTGGCAG CAACCAAGCT 20

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

GCCTCAGGCT GGGGCAGCAT T 21

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

GGTCACCTTC TGAGGGTGAA CTTGC 25

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

ACGACTCACT ATAAGCAGGA 20

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

AACAGCTATG ACCATCGTGG 20

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

ACGACTCACT ATGTGGAGAA 20

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

AACAGCTATG ACCCTGAGGA 20

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

GGAGCTGCCT GACGGCCAGG TCATC 25

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1599 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:115..744

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

GCGGCAGGCG CGGCAAATTA CGTTGCCGGA GCTGAACGGC GCGGCTGGTC TGAAGGCAAA 60

CAAGCGAGCG AGCGCGCGAT AGGGGCCGAG AGGACGCGCA GGTGGCGGCG TTGC ATG 117
Met
1

TCG CAC GGT CAC AGC CAC GGA ATG GGT GAC TGC CGC TGC GCC GCC GAA 165
Ser His Gly His Ser His Gly Met Gly Asp Cys Arg Cys Ala Ala Glu
5 10 15

CGG GAG GAG CCG CCC GAG CAG CAC GCC ATG GCT ACG CTG TAC CTG CGC 213
Arg Glu Glu Pro Pro Glu Gln His Ala Met Ala Thr Leu Tyr Leu Arg
20 25 30

ATC GAC CTG GAG CGG CTG CAA TGC CTT AAC GAG AGC CGC GAG GGC AGC 261
Ile Asp Leu Glu Arg Leu Gln Cys Leu Asn Glu Ser Arg Glu Gly Ser
35 40 45

GGC CGC GGC GTC TTC AAG CCG TGG GAG GAG CGG ACC GAC CGC TCC AAG 309
Gly Arg Gly Val Phe Lys Pro Trp Glu Glu Arg Thr Asp Arg Ser Lys
50 55 60 65

TTT GTT GAA AGT GAT GCA GAT GAA GAG CTT CTG TTT AAT ATT CCA TTT 357
Phe Val Glu Ser Asp Ala Asp Glu Glu Leu Leu Phe Asn Ile Pro Phe
70 75 80

ACG GGC AAT GTC AAG CTC AAA GGC ATC ATT ATA ATG GGA GAG GAT GAT 405
Thr Gly Asn Val Lys Leu Lys Gly Ile Ile Ile Met Gly Glu Asp Asp
85 90 95

GAC TCA CAC CCC TCT GAG ATG AGA CTG TAC AAG AAT ATT CCA CAG ATG 453
Asp Ser His Pro Ser Glu Met Arg Leu Tyr Lys Asn Ile Pro Gln Met
100 105 110

TCC TTT GAT GAT ACA GAA AGG GAG CCA GAT CAG ACC TTT AGT CTG AAC 501
Ser Phe Asp Asp Thr Glu Arg Glu Pro Asp Gln Thr Phe Ser Leu Asn

```
    115                 120                 125

CGG GAT CTT ACA GGA GAA TTA GAG TAT GCT ACA AAA ATT TCT CGT TTT      549
Arg Asp Leu Thr Gly Glu Leu Glu Tyr Ala Thr Lys Ile Ser Arg Phe
130                 135                 140                 145

TCA AAT GTC TAT CAT CTC TCA ATT CAT ATT TCA AAA AAC TTC GGA GCA      597
Ser Asn Val Tyr His Leu Ser Ile His Ile Ser Lys Asn Phe Gly Ala
                150                 155                 160

GAT ACG ACA AAG GTC TTT TAT ATT GGC CTG AGA GGA GAG TGG ACT GAG      645
Asp Thr Thr Lys Val Phe Tyr Ile Gly Leu Arg Gly Glu Trp Thr Glu
            165                 170                 175

CTT CGC CGA CAC GAG GTG ACC ATC TGC AAT TAC GAA GCA TCT GCC AAC      693
Leu Arg Arg His Glu Val Thr Ile Cys Asn Tyr Glu Ala Ser Ala Asn
        180                 185                 190

CCA GCA GAC CAT AGG GTC CAT CAG GTT ACC CCA CAG ACA CAC TTT ATT      741
Pro Ala Asp His Arg Val His Gln Val Thr Pro Gln Thr His Phe Ile
    195                 200                 205

TCC TAAGGGCTGG CCAAGGCTCC CATAGAGGCG CTGTGTCAGT GAAGATGTAC           794
Ser
210

GACTACCTGT TGGGAAGGAC AAAGGGATGA GGCTCCAGAG AGAGTTGGCT GCCACAGCTC    854

TGCCAAGCTT TGTCTTTGGG GCTTGCTGCA GAAACCTGGC CTACGGAAGA TACGACACCA    914

CTGGGAGGGT TGTGTAGGTG CCAGGGGACC ATCGTGGTTC TCTAGGGCGC TGTGGAAATT    974

GGGTCTTGGG CTGGGTGGCA TCTGGCAGTC ATGGGTAACA CTTGCTTTTC CAGTTAATGT   1034

GGCCATGTGA TTCCAAGTGT CATGTTGCTT TGTGGAAGAT TGTTGTGTGA CTTGTTTTTT   1094

TGATTTTGTA TTTGTTTTTT TAAAGGAAAC TATTTGTGGG CTATAGGAAA CTTTCTGATG   1154

CCTCCGGATT GTGTTAGTAG TAGCCATCAG GAGGGTCTCC AACTAAAACA CTTGTTCCTG   1214

CTTGCTCCTT TCCCCTCTCA TTGTTCAGCA TTCTTGTCAA GTTGCCCAGC TTGGAGTTGT   1274

CTGTCACGCA CATGTGTCCT GTGGTTATAG CTAGAAGGAC AGGAGTCTCC TGCTGATGCG   1334

TGATAGCTTA AGCTTGGGGA GAAGGTCTTT TCCACTGCCT AGCTAAGCAG TCTGGGGAGA   1394

GCATGGGGAT CATTTCTATG TGTGTGGGTA ATCTGGTCAG TAAGATTGAG ACTTAGTTAA   1454

GATTCCCCTT GGAAATTCCT TAATGTTTAT TAGCTTCTAA CTAGTGTTGT AAGTCCGATG   1514

CCAGAATTTG GAGATTTGAG TTCTTCTTTT CATGGCTTTT ATTCACTGTG ACTAATAAGC   1574

TTCCTAATAA ATCCTTGCCA GACTT                                         1599
```

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 210 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

Met Ser His Gly His Ser His Gly Met Gly Asp Cys Arg Cys Ala Ala
1 5 10 15

Glu Arg Glu Glu Pro Pro Glu Gln His Ala Met Ala Thr Leu Tyr Leu
20 25 30

Arg Ile Asp Leu Glu Arg Leu Gln Cys Leu Asn Glu Ser Arg Glu Gly
35 40 45

Ser Gly Arg Gly Val Phe Lys Pro Trp Glu Glu Arg Thr Asp Arg Ser
50 55 60

Lys Phe Val Glu Ser Asp Ala Asp Glu Glu Leu Leu Phe Asn Ile Pro
65 70 75 80

Phe Thr Gly Asn Val Lys Leu Lys Gly Ile Ile Ile Met Gly Glu Asp
85 90 95

Asp Asp Ser His Pro Ser Glu Met Arg Leu Tyr Lys Asn Ile Pro Gln
100 105 110

Met Ser Phe Asp Asp Thr Glu Arg Glu Pro Asp Gln Thr Phe Ser Leu
115 120 125

Asn Arg Asp Leu Thr Gly Glu Leu Glu Tyr Ala Thr Lys Ile Ser Arg
130 135 140

Phe Ser Asn Val Tyr His Leu Ser Ile His Ile Ser Lys Asn Phe Gly
145 150 155 160

Ala Asp Thr Thr Lys Val Phe Tyr Ile Gly Leu Arg Gly Glu Trp Thr
165 170 175

Glu Leu Arg Arg His Glu Val Thr Ile Cys Asn Tyr Glu Ala Ser Ala
180 185 190

Asn Pro Ala Asp His Arg Val His Gln Val Thr Pro Gln Thr His Phe
195 200 205

Ile Ser
210

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

CTGGCCTACG GAAGATACGA CAC 23

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

ACAATCCGGA GGCATCAGAA ACT 23

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

AGCCCCGGCC TCCTCGTCCT C 21

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

GGCGGCGGCA GCGGTTCTC 19

CLAIMS:

1. A method for identifying markers for a disease state, comprising the following steps:

a) providing a first set of peripheral blood mRNAs from one or more subjects known to exhibit said disease state and a second set of peripheral blood mRNAs from one or more normal subjects;

b) amplifying both sets of mRNAs to provide nucleic acid amplification products;

c) comparing said sets of amplification products; and

d) identifying those mRNAs that are differentially expressed between normal subjects and subjects exhibiting said disease state;

wherein a difference in quantity of expression of an mRNA is indicative of a disease marker.

2. The method of claim 1, further defined as comprising the step of using said mRNAs as templates for DNA synthesis in a reverse transcriptase reaction.

3. The method of claim 2, wherein random hexamers, arbitrarily chosen oligonucleotides, promiscuous oligonucleotide primers, anchoring primers or a combination of these are used as primers in the reverse transcriptase reaction.

4. The method of claim 1, wherein arbitrarily chosen oligonucleotides, promiscuous oligonucleotide primers, anchoring primers or a combination of these are used as primers in the amplification step.

5. The method of claim 1, wherein the disease state is metastatic or organ confined cancer, asthma, lupus erythematosis, rheumatoid arthritis, multiple sclerosis, myasthenia gravis, autoimmune thyroiditis, amyotrophic lateral sclerosis, interstitial cystitis or prostatitis.

6. The method of claim 5, wherein the disease state is metastatic prostate cancer.

7. The method of claim 5, wherein the disease state is metastatic breast cancer.

8. The method of claim 1, wherein said subjects are laboratory animals.

9. The method of claim 1, wherein said subjects are humans.

10. A method of detecting a metastatic cancer disease state in a subject, comprising the steps of:

a) detecting the quantity of expression of a metastatic cancer disease marker expressed in peripheral blood of said subject; and

b) comparing the quantity of expression of said marker in peripheral blood of said subject to the quantity of said marker expressed in peripheral blood of one or more normal subjects;

wherein a difference in quantity of expression of said marker in peripheral blood of said subject relative to quantity of expression of said marker in peripheral blood of said one or more normal individuals is indicative of a metastatic cancer disease state.

11. The method of claim 10, wherein said disease marker is an mRNA.

12. The method of claim 11, wherein said mRNA is amplified by an RNA polymerase reaction.

13. The method of claim 11, wherein said mRNA is amplified by reverse transcriptase polymerase chain reaction or ligase chain reaction.

14. The method of claim 10, wherein said detecting is by RNA fingerprinting, branched DNA or nuclease protection assay.

15. The method of claim 10, wherein said metastatic cancer disease state is metastatic prostate cancer.

16. The method of claim 10, wherein said metastatic cancer disease state is metastatic breast cancer.

17. The method of claim 11 in which said mRNA comprises one or more of the sequences or the complements of the sequences disclosed herein as Genebank Accession numbers D87451, T03013, X03558, M28130, Y00787, SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5 or SEQ ID NO:29.

18. The method of claim 10 in which said marker is a product of the interleukin 8 gene.

19. The method of claim 10, wherein said metastatic cancer disease marker is identified by the method of claim 1.

20. The method of claim 11, further defined as comprising the steps of

a) providing primers that selectively amplify at least a portion of said disease state marker;

b) amplifying said disease state marker with said primers to form nucleic acid amplification products;

c) detecting said nucleic acid amplification products; and

d) measuring the amount of said nucleic acid amplification products formed.

21. The method of claim 20 in which said primers are selected to produce an amplicon having a sequence of or complementary to a sequence of at least a 50 base contiguous segment of Genebank Accession numbers D87451, T03013, X03558, M28130, Y00787, SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:29.

22. The method of claim 21, wherein said amplicon is from about 50 to about 500 bases in length.

23. The method of claim 21, wherein said amplicon is from about 100 to about 415 bases in length.

24. The method of claim 10, wherein said metastatic cancer disease marker is a polypeptide.

25. The method of claim 24, wherein said polypeptide is encoded by a nucleic acid sequence comprising the sequence disclosed herein as Genebank Accession numbers D87451, T03013, X03558, M28130, Y00787, SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:29.

26. The method of claim 24, wherein said detection comprises antibody immunoreaction with said polypeptide.

27. The method of claim 26, wherein said detection comprises an ELISA, an immunoprecipitation, a radioimmunoassay, an immunohistochemical, Western blotting, dot blotting, or FACS analyses.

28. The method of claim 24, wherein said polypeptide is encoded by the IL-8 gene.

29. The method of claim 10 or claim 24, wherein said marker is a product of the IL-8 gene and wherein said comparison is between two alternatively spliced forms of an IL-8 gene product.

30. The method of claim 24, wherein the quantity of IL-8 polypeptide in peripheral blood is measured using an *in vitro* bioassay that detects at least one IL-8 mediated biological process.

31. The method of claim 29 wherein said markers comprise Genebank Accession # M28310, Y00787, SEQ ID NO:4 and SEQ ID NO:5.

32. A disease marker for prognosis or diagnosis of a disease condition, wherein said disease marker is identified by a process comprising:

a) providing a first set of peripheral blood mRNAs from one or more subjects known to exhibit said disease state and a second set of peripheral blood mRNAs from one or more normal subjects;

b) amplifying both sets of mRNAs to provide nucleic acid amplification products;

c) comparing said sets of amplification products; and

d) identifying those mRNAs that are differentially expressed between normal subjects and subjects exhibiting said disease state;

wherein a difference in quantity of expression of an mRNA is indicative of a disease marker.

33. The disease marker of claim 33, wherein the disease state is metastatic or organ confined cancer, asthma, lupus erythematosis, rheumatoid arthritis, multiple sclerosis, myasthenia gravis, autoimmune thyroiditis, amyotrophic lateral sclerosis, interstitial cystitis or prostatitis.

34. The method of claim 32, wherein the disease state is metastatic prostate cancer.

35. The method of claim 32, wherein the disease state is metastatic breast cancer.

36. The method of claim 32, wherein said subjects are laboratory animals.

37. The method of claim 32, wherein said subjects are humans.

38. A method of detecting prostate cancer in a biological sample, comprising:

(a) measuring the levels of IL-8 in combination with at least one prostate disease marker in said sample; and

(b) comparing said levels with corresponding levels obtained from reference populations of normal individuals, individuals with BPH and individuals with prostate cancer.

39. The method of claim 38 in which said prostate disease marker is selected from a group consisting of: total prostate specific antigen (PSA); prostate specific membrane antigen (PSMA=Folic Acid Hydrolase); prostate acid phosphatase (PAP); prostatic secretory proteins

(PSP_{94}); human kallekrein 2 (HK2); and the ratio of the concentrations of free and bound forms of PSA (f/t PSA).

40. The method of claim 38 in which the biological sample comprises peripheral human blood.

41. The method of claim 38 wherein the level of IL-8 in a biological sample is measured using at least one antibody that binds to at least one IL-8 gene product.

42. The method of claim 41 wherein the level of IL-8 gene product bound to antibody is measured by ELISA.

43. The method of claim 38 wherein the level of IL-8 in a biological sample is measured using at least one oligonucleotide probe that binds to at least one IL-8 messenger RNA (mRNA).

44. The method of claim 43 wherein the IL-8 mRNA is alternatively spliced to include intron 3.

45. The method of claim 43 wherein the level of oligonucleotide probe bound to IL-8 mRNA is measured by nuclease protection assay.

46. The method of claim 43 wherein the level of oligonucleotide probe bound to IL-8 mRNA is measured by RT-PCR™.

47. The method of claim 43 wherein the level of oligonucleotide probe bound to IL-8 mRNA is measured by ligase chain reaction.

48. The method of claim 43 wherein the level of oligonucleotide probe bound to IL-8 mRNA is measured by PCR™.

49. The method of claim 40 wherein the level of IL-8 in a biological sample is measured using an *in vitro* bioassay that detects at least one IL-8 mediated biological process.

50. The method of claim 44 wherein the level of IL-8 in a biological sample is measured using at least one molecule that binds to an IL-8 gene product, wherein said molecule is selected from a group consisting of: an IL-8 binding protein; and an IL-8 receptor protein.

51. The method of claim 48 wherein the level of prostate disease marker in a biological sample is measured using at least one antibody that binds to at least one prostate disease marker protein.

52. The method of claim 51 wherein the level of prostate disease marker protein bound to antibody is measured by ELISA.

53. The method of claim 39 wherein the level of prostate disease marker in a biological sample is measured using at least one oligonucleotide probe that binds to at least one prostate disease marker messenger RNA (mRNA).

54. The method of claim 43 wherein the level of oligonucleotide probe bound to prostate disease marker mRNA is measured by nuclease protection assay.

55. The method of claim 43 wherein the level of oligonucleotide probe bound to prostate disease marker mRNA is measured by RT-PCR™.

56. The method of claim 43 wherein the level of oligonucleotide probe bound to prostate disease marker mRNA is measured by ligase chain reaction.

57. The method of claim 43 wherein the level of oligonucleotide probe bound to prostate disease marker mRNA is measured by PCR™.

58. A method of differentially diagnosing prostate cancer and benign prostatic hyperplasia, comprising the step of measuring the levels of IL-8 in combination with at least one prostate disease marker in a biological sample.

59. The method of claim 58 in which said prostate disease marker is selected from a group consisting of: total prostate specific antigen (PSA), prostate specific membrane antigen (PSMA=Folic Acid Hydrolase), prostate acid phosphatase (PAP), prostatic secretory proteins (PSP_{94}), human kallekrein 2 (HK2), and the ratio of the concentrations of free and bound forms of PSA (f/t PSA).

60. The method of claim 59 in which said biological sample consists of peripheral human blood.

61. A kit for use in detecting a human disease, comprising:

(a) a pair of primers for amplifying a disease state marker consisting of a nucleic acid; and

(b) containers for each of said primers.

62. A kit according to claim 61 in which the pair of primers is selected to amplify a nucleic acid marker for metastatic human cancer.

63. A kit according to claim 62 in which the pair of primers is selected to amplify a nucleic acid having a sequence comprising at least a 50 base segment of Genebank Accession numbers D87451, T03013, X03558, M28130, Y00787, SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:29.

64. A kit according to claim 62, comprising:

(a) a pair of primers selected to amplify a nucleic acid sequence comprising SEQ ID NO:4 or Genebank Accession # Y00787; and

(b) a pair of primers selected to amplify a nucleic acid sequence comprising SEQ ID NO:5 or Genebank Accession # M28130.

65. A kit for use in diagnosing metastatic cancer in a biological sample, comprising:

(a) an antibody which binds with high specificity to a polypeptide having an amino acid sequence encoded by a nucleic acid sequence comprising Genebank Accession numbers D87451, T03013, X03558, M28130, Y00787, SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:29.

(b) a container for said antibody.

66. A kit according to claim 65, further defined as comprising:

(a) an antibody that binds with high specificity to a soluble IL-8 gene product;

(b) an antibody that binds with high specificity to a membrane bound IL-8 gene product; and

(c) a container for each antibody.

67. A kit according to claim 65, wherein said metastatic cancer is metastatic prostate cancer.

68. A kit according to claim 65, wherein said metastatic cancer is metastatic breast cancer.

69. A kit for detecting or differentially diagnosing human prostate cancer, comprising:

(a) at least one detection agent for measuring the levels of IL-8 in a biological sample;

(b) at least one detection agent for measuring the levels of at least one prostate disease marker in said biological sample; and

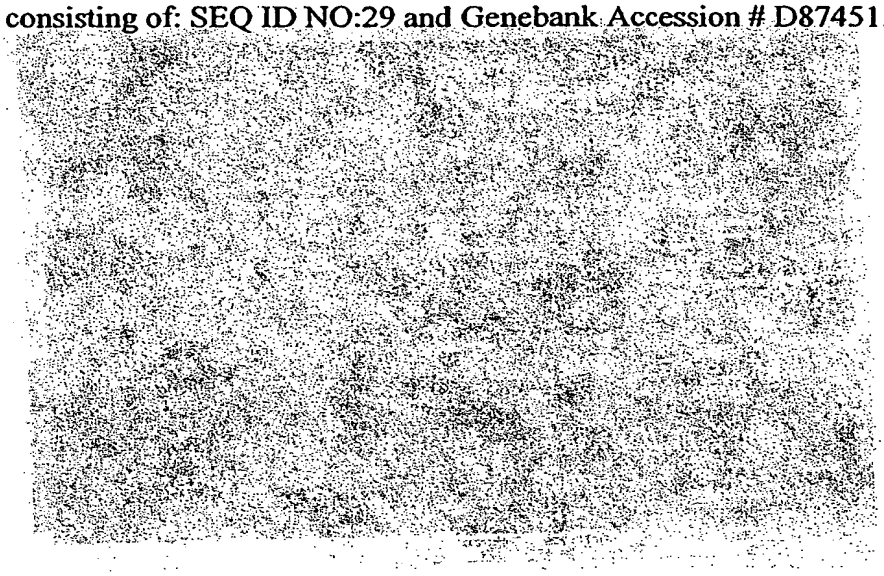
(c) containers for each of said detection agents.

70. The kit of claim 69 in which said prostate disease marker is selected from a group consisting of: total prostate specific antigen (PSA), prostate specific membrane antigen (PSMA=Folic Acid Hydrolase), prostate acid phosphatase (PAP), prostatic secretory proteins

(PSP_{94}), human kallekrein 2 (HK2), and the ratio of the concentrations of free and bound forms of PSA (f/t PSA).

71. The kit of claim 70 in which said detection agents are selected from a group consisting of: polyclonal antibodies; monoclonal antibodies; oligonucleotides; paired oligonucleotides designed to bind to opposite strands of a double-stranded DNA molecule; and at least one molecule that binds to an IL-8 gene product.

72. The method of claim 16 in which said breast cancer marker is selected from a group consisting of: SEQ ID NO:29 and Genebank Accession # D87451.



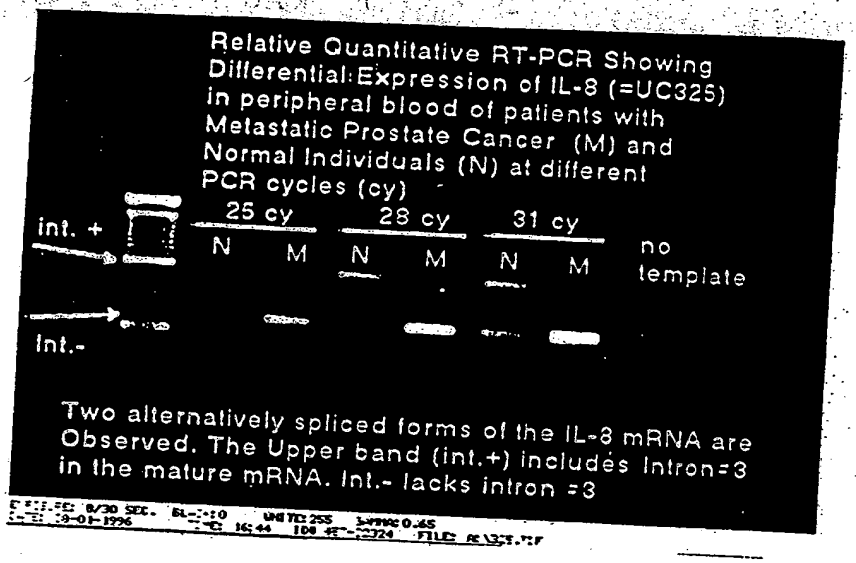

Relative Quantitative RT-PCR Showing Differential Expression of IL-8 (=UC325) in peripheral blood of patients with Metastatic Prostate Cancer (M) and Normal Individuals (N) at different PCR cycles (cy)
25 cy
28 cy
31 cy
N
M
N
M
N
M
no template
int. +
int.-
Two alternatively spliced forms of the IL-8 mRNA are Observed. The Upper band (int.+) includes Intron=3 in the mature mRNA. Int.- lacks intron =3

FIG. 1A

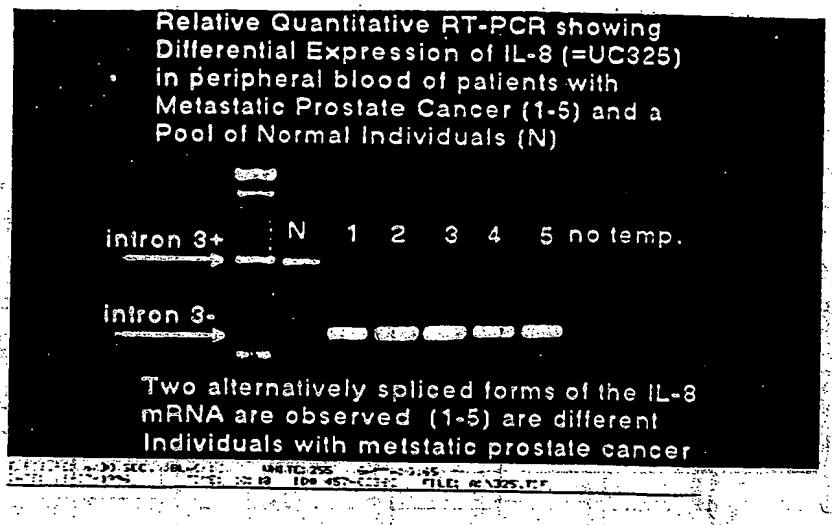

Relative Quantitative RT-PCR showing
Differential Expression of IL-8 (=UC325)
in peripheral blood of patients with
Metastatic Prostate Cancer (1-5) and a
Pool of Normal Individuals (N)
intron 3+
N 1 2 3 4 5 no temp.
intron 3-
Two alternatively spliced forms of the IL-8
mRNA are observed (1-5) are different
Individuals with metstatic prostate cancer

FIG. 1B

Figure 2

Ability of Total PSA (ng/ml) to Distinguish BPH and Stages A, B, & C Prostate Cancer (n = 142)

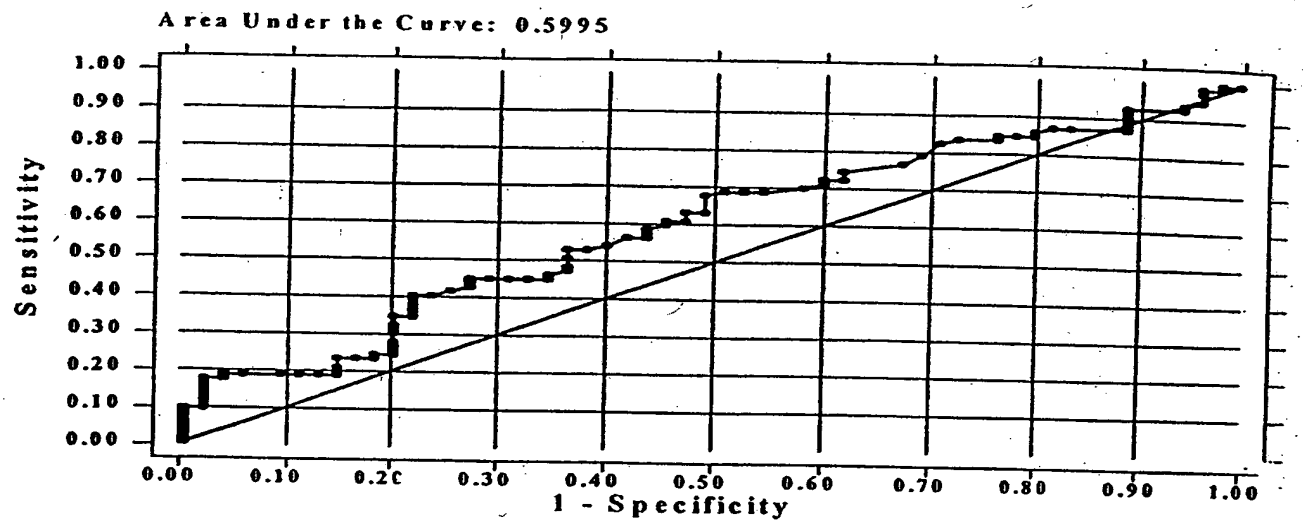


Figure 3

Ability of Corrected Free/Total PSA Ratio to Distinguish BPH and Stages A, B, & C Prostate Cancer (n = 142)

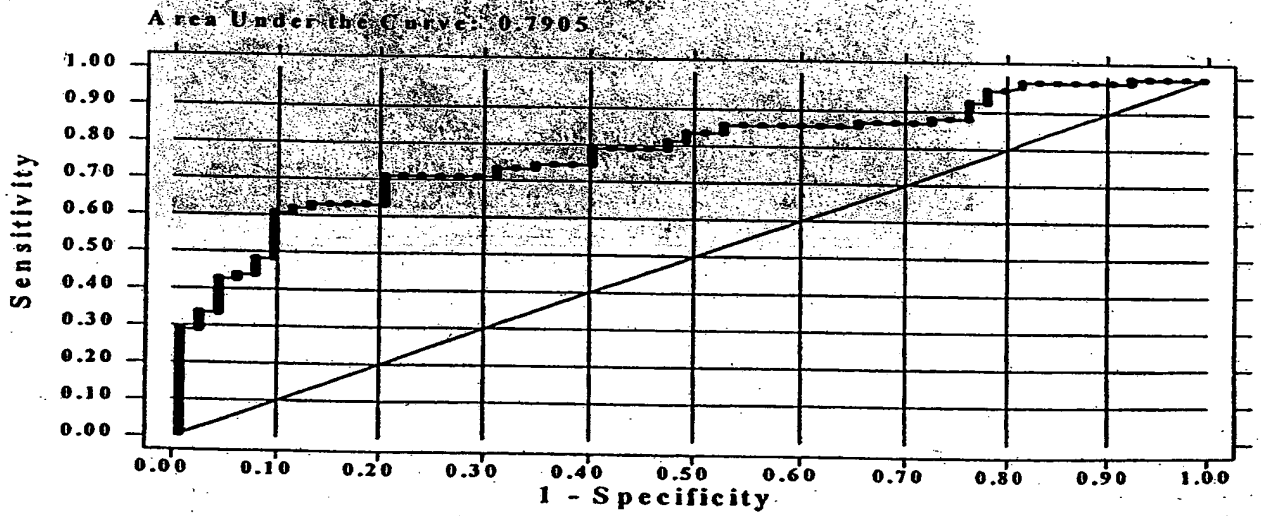


Figure 4

Ability of UC325 (pg/ml) to Distinguish BPH and Stages A, B, & C Prostate Cancer (n = 142)

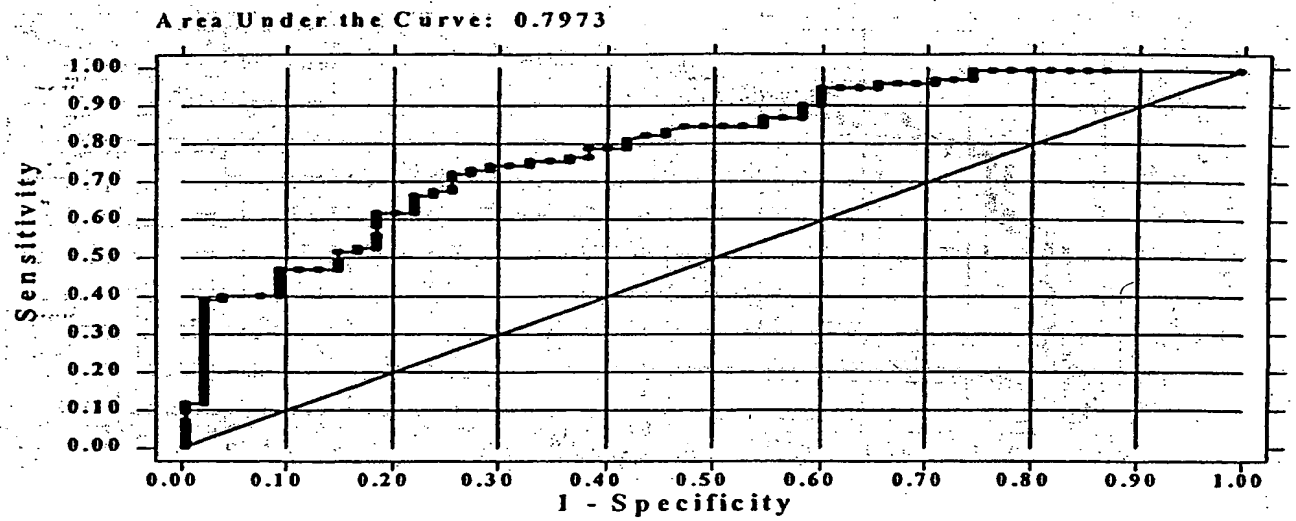


Figure 5

Ability of UC325 (pg/ml) & T-PSA (ng/ml) to Distinguish BPH and Stages A, B, & C Prostate Cancer (n = 142)

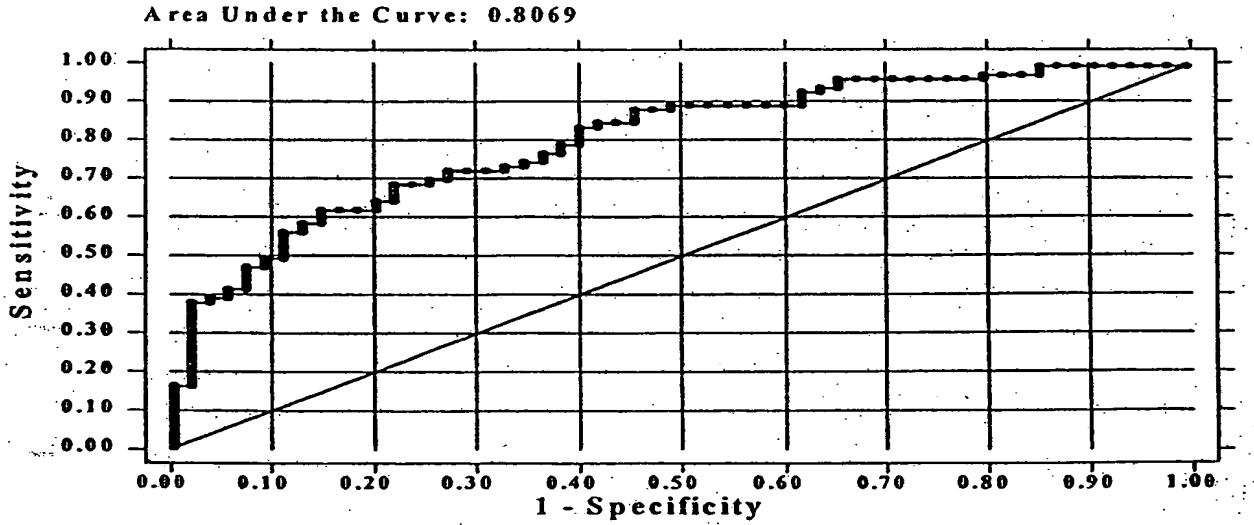
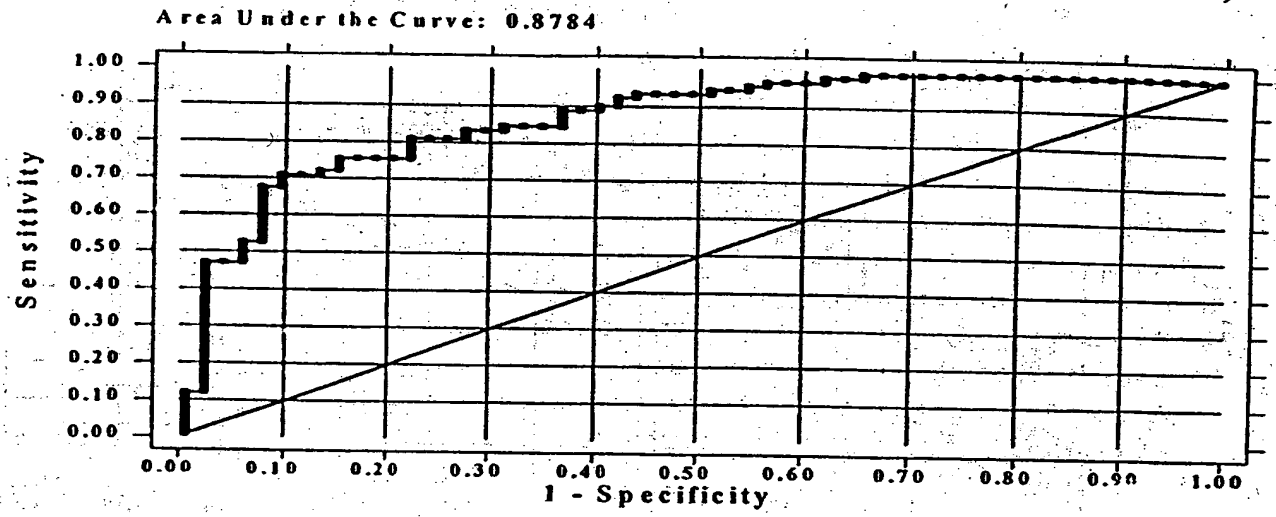


Figure 6

Ability of UC325 (pg/ml) & f/t PSA Ratio to Distinguish BPH and Stages A, B, & C Prostate Cancer (n = 142)



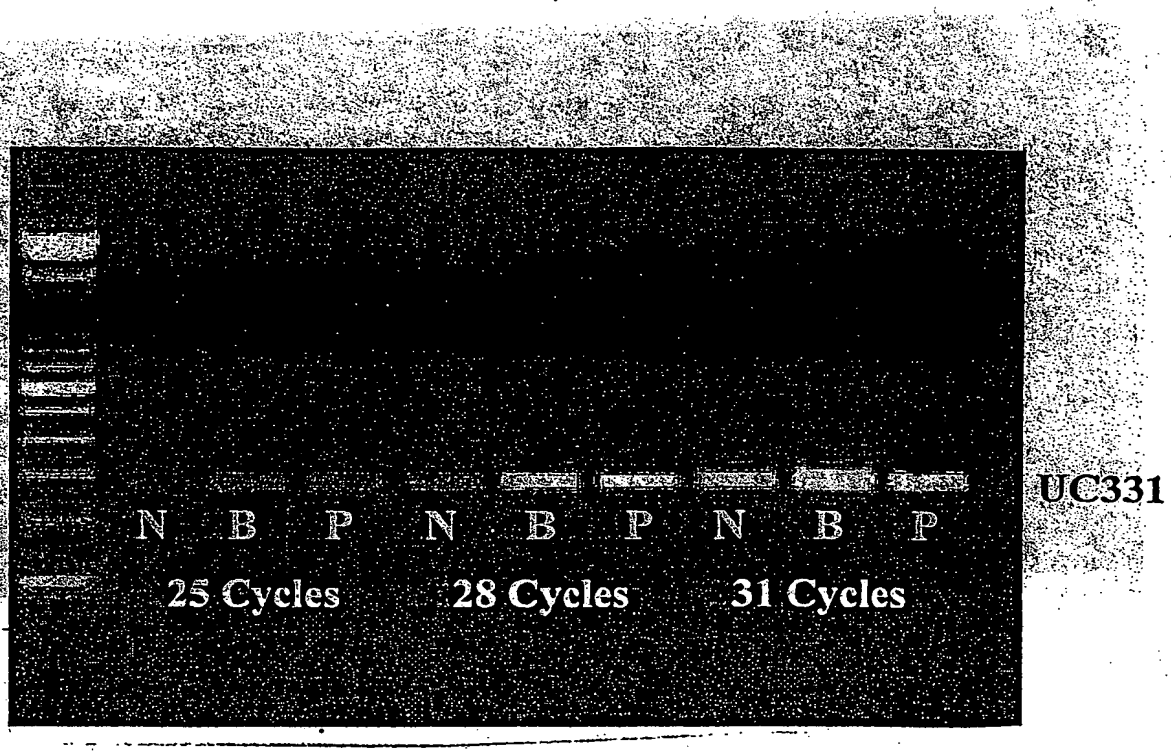


FIG. 7

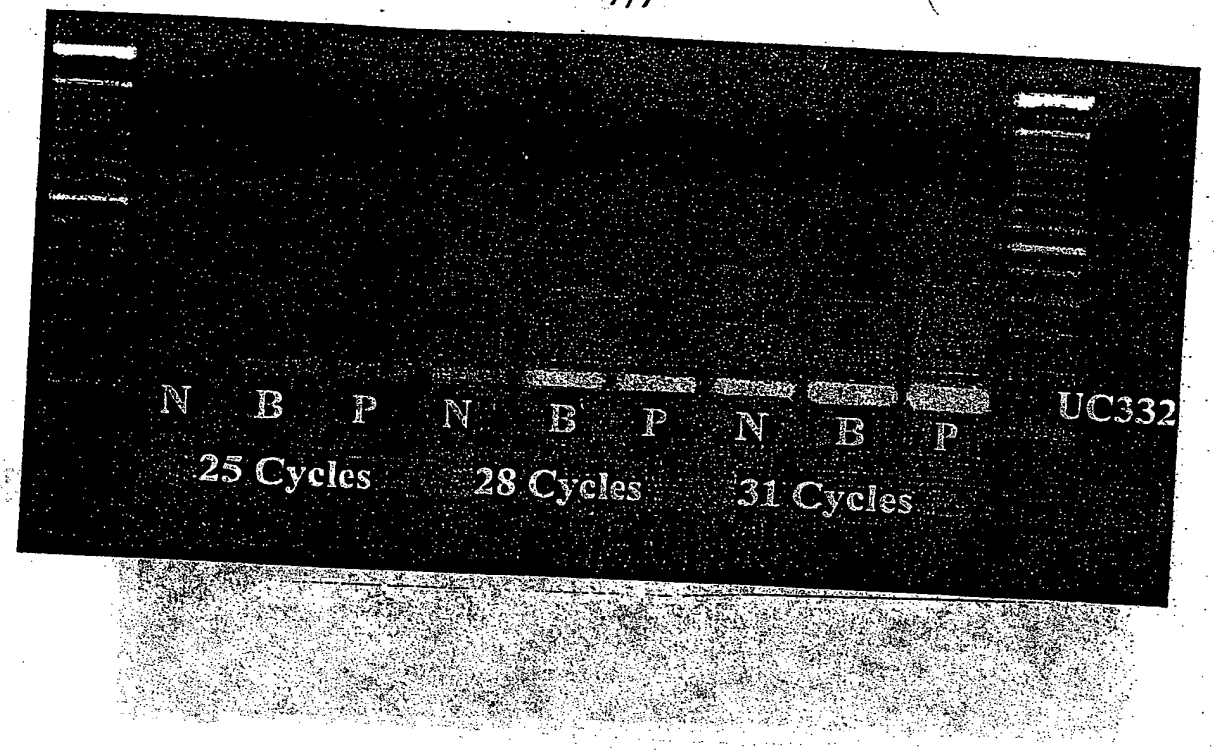


FIG. 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/22105

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C12Q 1/68; C12P 19/34; C07H 21/04; C07K 14/435; G01N 33/53
US CL :435/ 6; 7.1, 91.2, 91.21; 536/ 23.5, 24.31, 24.33; 530/350, 388.8

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/ 6; 7.1, 91.2, 91.21; 536/ 23.5, 24.31, 24.33; 530/350, 388.8

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

NONE

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y,	US 5,539,096 A (BABAI et al) 23 July 1996, col. 2 and 5.	1-72
Y, P	US 5,677,125 A (HOLT et al) 14 October 1997, col. 7, 25, 26.	1-72
Y	US 5,236,844 A (BASSET et al) 17 August 1993, col. 5 and 7.	1-72
Y	US 5,459,037 A (SUTCLIFFE et al) 17 October 1995, col. 14-16.	1-72
Y	IVANOVA et al. Identification of differentially expressed genes by restriction endonuclease-based gene expression fingerprinting. Nucleic Acids Research. 1995, Vol. 23, No. 15, pages 2954-2958, especially pages 2954-2955.	14

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

23 FEBRUARY 1998

Date of mailing of the international search report

20 MAR 1998

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer

CARLA MYERS

Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/22105

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	SCHMID et al. Induction of mRNA for a serine protease and a B-thromboglobulin-like protein in mitogen-stimulated human leukocytes. The Journal of Immunology. 01 July 1987, Vol. 139, No. 1, pages 250-256, especially page 252.	17, 21, 24, 28-31, 63-65
Y,P	GREENE et al. Correlation of metastasis-related gene expression with metastatic potential in human prostate carcinoma cells implanted in nude mice using an in situ messenger RNA hybridization technique. American Journal of Pathology. May 1997, Vol. 150, No. 5, pages 1571-1582, especially page 1580.	38-60, 69-71
Y, P	Database GenBank on STN, GenBank Accession No. D87451, NOMURA, N. 'Human mRNA for KIAA0262 gene, complete cds,' 10 July 1997.	72
Y	MATSUSHIMA et al. Molecular cloning of a human monocyte-derived neutrophil chemotactic factor (MDNCF) and the induction of MDNCF mRNA by interleukin 1 and tumor necrosis factor. Journal of Experimental Medicine. June 1988, Vol. 167, pages 1883-1893, especially page 1886.	17, 21, 24, 28-31, 63-65

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/22105

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS; Dialog: Medline, CA, Derwent Patents, Biosis, Embase, GenBank
search terms: SEQ ID NO: 1-5, 29; mRNA, differential display or expression, cancer or tumor or metastasis